

CURRENT ISSUE

[Table of Contents](#)



○ THIS WEEK

- [Editorials](#)
- [World View](#)
- [Research Highlights](#)
- [Seven Days](#)

○ NEWS IN FOCUS

- [News](#)
- [Feature](#)

○ COMMENT

- [Comment](#)
- [Books and Arts](#)
- [Correspondence](#)
- [Obituary](#)

○ CAREERS

- [Feature](#)
- [Column](#)

▪ [Futures](#)

○ SPECIALS

- [Outlook](#)

○ RESEARCH

- [News & Views](#)
- [Reviews](#)
- [Article](#)
- [Letters](#)

- [Previous issue](#)

THIS WEEK

EDITORIALS

- **[The morning after](#)**

President Barack Obama's stance on an emergency contraceptive betrays his promised principles of scientific integrity and sets a troubling precedent for political interference in 'inconvenient' science.

- **[Defend the Amazon](#)**

Brazilian lawmakers should not weaken their stance on deforestation to appease landowners.

- **[Ten for 2011](#)**

As the year ends, Nature highlights individuals who rose to prominence — or fell from grace.

WORLD VIEW

- **[Particle physics is at a turning point](#)**

The discovery of the Higgs boson will complete the standard model — but it could also point the way to a deeper understanding, says Gordon Kane.

RESEARCH HIGHLIGHTS

- **[Genetics: Case of a missing cluster](#)**
- **[Planetary science: Mercury's plasma belt](#)**
- **[Zoology: Birds keep up with the Joneses](#)**
- **[Cell biology: The first microtubules](#)**
- **[Microbiology: Splitting a bacterial magnet](#)**
- **[Mathematics: What da Vinci saw in trees](#)**
- **[Animal behaviour: Frogs croak about their pad](#)**
- **[Cancer biology: Sisterhood of lymphoma](#)**
- **[Evolutionary biology: Modern humans smell good](#)**

- 

[Networks: Control of the super-corporations](#)

SEVEN DAYS

- **[Seven days: 16–22 December 2011](#)**

The week in science: Cornell to build US\$2-billion science campus in New York; Kepler finds a twin Earth; and Fukushima is declared to be in cold shutdown.

NEWS IN FOCUS



Fears grow over lab-bred flu

Scientists call for stricter biosafety measures for dangerous avian-influenza variants.

Declan Butler



Last-minute wins for US science

Bill tops up health, energy and translational-science spending.

Ivan Semeniuk & Susan Young



Chimp research under scrutiny

US-government-funded studies to be evaluated under stricter criteria.

Meredith Wadman



Rules on integrity signal tighter stance

Many US agencies now have policies banning political manipulation of research — but ambiguities remain.

○ Eugenie Samuel Reich

- **Correction**



365 days: 2011 in review

From neutrinos to stem cells: a round-up of the year in research and science policy.

○ Richard Van Noorden



365 days: Images of the year

Flying rhinos and furious rats vie with graphene knots and space technology in 2011's most striking pictures.

○ Daniel Cressey

FEATURE

- **365 days: *Nature's* 10**

Ten people who mattered this year.



Declan Butler, Ewen Callaway, Erika Check Hayden, David Cyranoski, Eric Hand **+ et al**

COMMENT

- **Crisis response: The new history**

The past year has shown how prone a highly networked society is to abrupt change. The future of our complex world, says Philip Ball, depends on becoming resilient to shocks.

- **Science publishing: The paper is not sacred**

Peer review continues long after a paper is published, and that analysis should become part of the scientific record, say Adam Marcus and Ivan Oransky.

BOOKS AND ARTS

- **Technology: Rise of the e-book**

Carl Zimmer charts the boom in electronic publishing and what that spells for wood pulp and ink.

- **Food science: With pipette and ladle**

From stretchy ice cream to wire-brushed crackling, Harold McGee digests an eclectic modernist menu.

- Review of *The Kitchen as Laboratory: Reflections on the Science of Food and Cooking*
César Vega, Job Ubbink & Erik Van Der Linden

- **Q&A: The snowflake designer**

For the past decade, physicist Kenneth Libbrecht has been studying how ice crystals form, taking thousands of photographs of their intricate structures. He describes how he grows snowflakes in his lab at the California Institute of Technology in Pasadena, and never tires of tracking the real thing in the far north.

- **In retrospect: On the Six-Cornered Snowflake**

Philip Ball celebrates the fourth centenary of Johannes Kepler's ice-crystal analysis.

- Review of *De nive sexangula*
Johannes Kepler

CORRESPONDENCE

- **Art history: No refraction in Leonardo's orb**

- André J. Noest & Martin Kemp

- **Scavenger turned predator: European vultures' altered behaviour**

- Antoni Margalida, David Campión & José A. Donázar

- **Founding fathers: The cell was defined 150 years ago**

- U. Kutschera

OBITUARY

- **Lynn Margulis (1938–2011)**

Biologist who revolutionized our view of early cell evolution.

- James A. Lake

CAREERS

FEATURES

- **Awards: Conscientious counsellors**
Nature's mentoring awards honour three scientists in France.
- Declan Butler

COLUMNS

- **Testing the waters**
Postdoc committees can give insight into industry career paths, argue Christopher Tsang and Michael Fisher.
- Christopher Tsang & Michael Fisher

naturejobs job listings and advertising features

FUTURES

- **Tea with Jillian**
Robotic reflections.
- Brenda Cooper

SPECIALS

OUTLOOK: TRADITIONAL ASIAN MEDICINE

Produced with support from:

Outlook

- **Traditional Asian medicine**
○ Michelle Grayson
- **TCM: Made in China**
○ Felix Cheung
- **Convergence: Where West meets East**
○ Peng Tian
- **Perspective: All systems go**
○ Jan van der Greef
- **Microbiome: That healthy gut feeling**
○ James Mitchell Crow
- **Modernization: One step at a time**
○ Zhiguo Xu
- **Patents: Protecting China's national treasure**

- Jessie Jiang
-

- **Modern TCM: Enter the clinic**

- Felix Cheung
-

- **Japan: Will the sun set on Kampo?**

- Ichiko Fuyuno
-

- **Perspective: Herbal dangers**

- Masatomo Sakurai
-

- **Regulations: Herbal medicine rule book**

- Natasha Gilbert
-

- **Perspective: The clinical trial barriers**

- Liang Liu, Elaine Lai-Han Leung & Xiaoying Tian
-

- **Biodiversity: Endangered and in demand**

- Duncan Graham-Rowe

RESEARCH

NEWS & VIEWS

- **Planetary science: The ultimate fate of planets**

- Eliza M. R. Kempton

See also

- [Letter by Charpinet et al.](#)

- **Ecology: Bleak future for amphibians**

- Ross A. Alford

See also

- [Letter by Hof et al.](#)

- **50 & 100 years ago**

- **Psychology: Who needs a leader?**

- Sadaf Shadan

- **Atomic physics: When ultracold is not cold enough**

- Gretchen K. Campbell

See also

- [Letter by Bakr et al.](#)

- **Condensed-matter physics: A fresh twist on shrinking materials**

- J. Paul Attfield

- **Physiology: On time metabolism**

- Joseph Bass

See also

- [Letter by Lamia et al.](#)

- **2011 Editors' choice**

REVIEWS

- **Rethinking amide bond synthesis**
- Vijaya R. Pattabiraman & Jeffrey W. Bode
- **Cancer immunotherapy comes of age**
- Ira Mellman, George Coukos & Glenn Dranoff

ARTICLE

- **DNA-binding factors shape the mouse methylome at distal regulatory regions**
- Michael B. Stadler, Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert **+ et al**

LETTERS

- **A compact system of small planets around a former red-giant star**
- S. Charpinet, G. Fontaine, P. Brassard, E. M. Green, V. Van Grootel **+ et al**
- See also
 - **News & Views by Kempton**
- **Orbital excitation blockade and algorithmic cooling in quantum gases**
- Waseem S. Bakr, Philipp M. Preiss, M. Eric Tai, Ruichao Ma, Jonathan Simon **+ et al**
- See also
 - **News & Views by Campbell**
- **Tuning charge transport in solution-sheared organic semiconductors using lattice strain**
- Gaurav Giri,
- Eric Verploegen, Stefan C. B. Mannsfeld, Sule Atahan-Evrenk, Do Hwan Kim **+ et al**
- **Forcing of wet phases in southeast Africa over the past 17,000 years**
- Enno Scheuß, Holger Kuhlmann, Gesine Mollenhauer, Matthias Prange & Jürgen Pätzold
- **Lowland–upland migration of sauropod dinosaurs during the Late Jurassic epoch**
- Henry C. Fricke, Justin Henceroth & Marie E. Hoerner
- **Additive threats from pathogens, climate and land-use change for global amphibian diversity**
- Christian Hof, Miguel B. Araújo, Walter Jetz & Carsten Rahbek
- See also
 - **News & Views by Alford**
- **The *Medicago* genome provides insight into the evolution of rhizobial symbioses**
- Nevin D. Young, Frédéric Debellé, Giles E. D. Oldroyd, Rene Geurts, Steven B. Cannon **+ et al**
- **Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour**
- Howard C. Chang, Jennifer Paek & Dennis H. Kim
- **Adherens junction protein nectin-4 is the epithelial receptor for measles virus**
- Michael D. Mühlebach, Mathieu Mateo, Patrick L. Sinn, Steffen Prüfer, Katharina M. Uhlig **+ et al**
- **Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum***
- Cécile Crosnier, Leyla Y. Bustamante, S. Josefin Bartholdson, Amy K. Bei, Michel Theron **+ et al**
- **Response to self antigen imprints regulatory memory in tissues**

- Michael D. Rosenblum, Iris K. Gratz, Jonathan S. Paw, Karen Lee, Ann Marshak-Rothstein [+ et al](#)
- [Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease](#)
- Philipp Koch, Peter Breuer, Michael Peitz, Johannes Jungverdorben, Jaideep Kesavan [+ et al](#)
- [Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson’s disease](#)
- Sonja Kriks, Jae-Won Shim, Jinghua Piao, Yosif M. Ganat, Dustin R. Wakeman [+ et al](#)
- [Cryptochromes mediate rhythmic repression of the glucocorticoid receptor](#)
- Katja A. Lamia, Stephanie J. Papp, Ruth T. Yu, Grant D. Barish, N. Henriette Uhlenhaut [+ et al](#)

See also

- [News & Views by Bass](#)
- [GlcNAcylation of histone H2B facilitates its monoubiquitination](#)
- Ryoji Fujiki,
- Waka Hashiba, Hiroki Sekine, Atsushi Yokoyama, Toshihiro Chikanishi [+ et al](#)
- [An equilibrium-dependent retroviral mRNA switch regulates translational recoding](#)
- Brian Houck-Loomis, Michael A. Durney, Carolina Salguero, Neelaabh Shankar, Julia M. Nagle [+ et al](#)
- [Structures of the multidrug exporter AcrB reveal a proximal multisite drug-binding pocket](#)
- Ryosuke Nakashima, Keisuke Sakurai, Seiji Yamasaki, Kunihiro Nishino & Akihito Yamaguchi
- [Intermediates in the transformation of phosphonates to phosphate by bacteria](#)
- Siddhesh S. Kamat, Howard J. Williams & Frank M. Raushel

THIS WEEK

EDITORIALS

NATURE'S TEN How we chose the heroes and villain of the past year **p.414**

WORLD VIEW We may have found the Higgs boson, what now for physics? **p.415**

KISS ME Teeny tiny frog shows itself (just) on forest floor **p.419**



The morning after

President Barack Obama's stance on an emergency contraceptive betrays his promised principles of scientific integrity and sets a troubling precedent for political interference in 'inconvenient' science.

In the opening months of his administration, US President Barack Obama gathered scientists in the grand East Room of the White House to promise that things would be different on his watch. The president was there to sign a new memo on scientific integrity that directed agency heads not to meddle in the decisions of their scientific staff.

In his speech, Obama pledged that his ambition was “about letting scientists like those here today do their jobs, free from manipulation or coercion, and listening to what they tell us, even when it's inconvenient — especially when it's inconvenient”.

Yet events earlier this month saw Obama demonstrate a blatant betrayal of these principles, when he defended a decision by Kathleen Sebelius, US secretary of health and human services, to override a drug approval by the Food and Drug Administration (FDA).

Scientific staff at the drug-regulatory agency, including paediatricians, obstetricians and gynaecologists, reached a key determination earlier this year involving Plan B One-Step (levonorgestrel), a single-dose ‘morning-after’ contraceptive pill. They concluded that it is safe and effective for girls younger than 17 years of age without the currently required prescription — and that these girls could use it correctly without a physician's help. (Older women can already get the contraceptive over the counter.) The FDA experts came to their decision after conducting the usual extensive review, including a comprehensive survey of the literature and data from the pill's manufacturer, Teva, headquartered in Petah Tikva, Israel. These data included age-specific studies designed specifically to address the agency's standards for non-prescription drugs. The scientific staff concluded that, as FDA commissioner Margaret Hamburg put it earlier this month: “There is adequate and reasonable, well-supported, and science-based evidence that Plan B One-Step is safe and effective and should be approved for nonprescription use for all females of child-bearing potential.”

So much for science. One day after Sebelius announced that she was overruling the agency and denying over-the-counter access to the morning-after pill to females under 17, Obama, questioned by a reporter, said that he agreed with her decision, and added that, “as the father of two daughters” he lauded the application of “common sense”.

He said: “The reason Kathleen made this decision was she could not be confident that a 10-year-old or an 11-year-old go into a drug-store, should be able — alongside bubble gum or batteries — be able to buy a medication that potentially, if not used properly, could end up having an adverse effect. And I think most parents would probably feel the same way.”

Where to start? Never mind that the most powerful man in the world didn't bother to get the basic facts right: the contraceptive, even for non-prescription use, is stored behind the pharmacy counter and not “beside bubble gum and batteries”. And never

mind the throwback paternalism implied by elevating his own opinion as “the father of two daughters”. What is most infuriating about Obama's response is the casual jettisoning of careful and

“What is most infuriating is the casual jettisoning of careful evidence on this contraceptive's safety.”

convincing evidence on this contraceptive's safety and effectiveness, even for young adolescents, and their ability to comprehend the labelling. (It essentially says: “Take this one pill, now.”) With it went his own promise to the thousands of scientists whose hard work and science he had pledged to respect, “especially when it's inconvenient”. It certainly is inconvenient, on the cusp of

an election year, in what is at heart a deeply conservative country, to acknowledge that young adolescents can and do have sex, and that they may not have thought out the potential consequences in advance. So inconvenient, apparently, that the work of the scientists, who spent long hours weighing risks and benefits for the public good, must be thrown under a bus.

Most troubling of all is that Obama has set a precedent for overriding science. Sebelius's trampling of a drug decision by the FDA is the first in living memory. If it is acceptable for her to override the agency for this year's reasons of political convenience — or, for that matter, for reasons of heartfelt belief — what politically loaded drugs will next come under the axe?

If there is a silver lining here, it is that some agencies seem to be listening to what the president says rather than following what he does. Several are working hard to put in place the integrity policies he mandated nearly three years ago (see page 425). The Department of Health and Human Services has not revealed where it stands on crafting such a policy of its own. ■

Defend the Amazon

Brazilian lawmakers should not weaken their stance on deforestation to appease landowners.

In the Amazon rainforest, there is a fine balance between efforts to prevent deforestation and the desire to clear the trees for agriculture. It says much about this tension that spikes in forest loss in Brazil earlier this year have been attributed to nothing more tangible than a perceived shift in where the country's politicians intend to draw the line between the two. After five years of stunning success in combating deforestation, the Brazilian authorities had started to ponder how to respond to the growing clamour from landowners and farmers

in the region who were being prevented from clearing land to cash in on record prices for commodities such as soya and beef. That sense of a weakening political resolve to protect the rainforest was enough of an incentive for some to begin clearing it again.

Those wielding the chainsaws and driving the bulldozers may have judged the climate correctly: lobbying and political discussions have ultimately produced a controversial bill to change how the Amazon is protected in Brazil. Final voting on amendments to the forest code was due in the Brazilian parliament this month, but has been postponed until March. Supporters and critics of the legislation are gearing up for a final push, and both sides are putting pressure on President Dilma Rousseff, who has the power to veto sections of the bill, or the bill itself.

The proposed forest code would update regulations dating back to 1965 on how private landowners must preserve native forest. Under the existing rules, they must maintain forest on 80% of their land, and those who have cleared illegally must restore to that level. The proposed change would remove the 80% obligation from small landowners, and grant an effective amnesty to those who illegally deforested their land before 2008, removing the threat of legal sanctions and fines for those who agree to reforest.

The government has said that the legislation improved as it moved through the Senate, but there are certainly problems remaining with the proposals. The bill undermines the old code's base in ecology, in that it would loosen restrictions on cutting trees in areas around rivers and on steep hills — rules that are intended to protect river health and prevent soil that is normally protected by vegetation being washed into waterways. This is just bad policy.

Through its exemption for small landowners, the revised code will legalize massive new destruction of forest — about 220,000 square

kilometres according to an analysis from researchers at the University of São Paulo — and there is reason for concern that the amnesty being granted could encourage further illegal deforestation, by giving landowners the impression that the government doesn't have what it takes to truly enforce the law. Furthermore, the requirements for reforestation by landowners who have broken the law are too weak.

"The proposed new forest code grants an effective amnesty to those who illegally deforested their land before 2008."

Scientists and environmentalists should continue to press for changes to the legislation in these and other areas. But they should also acknowledge that there are problems with the existing system — notably, that those landowners who have abided by the law have done so without reward, despite the promise of carbon payments down the road in exchange for the protection and stewardship of standing forest.

The Brazilian government maintains that it will meet its pledge to reduce deforestation by 80% by 2020, which was set by former president Luiz Inácio Lula da Silva at the Copenhagen climate conference in 2009, and that it is on target to do so.

Deforestation currently accounts for about 15% of global greenhouse-gas emissions, and some 75% of Brazil's. Meeting the pledge would be a massive achievement, and one that would allow Brazil to claim a place at the front of the global pack in terms of reducing carbon emissions and protecting biodiversity. But the government cannot get there on its own. It needs its policy to have broad support, or at least command respect, from people on the ground on all sides of the debate. And in this sense, the real danger isn't the new forest bill itself, but the sentiment of relaxed protection for the Amazon that seems to be behind it. ■

Ten for 2011

As the year ends, Nature highlights individuals who rose to prominence — or fell from grace.

Rebellion. Tragedy. Breakthrough. Crime. These are just a few of the major events that had an impact on science this year. Revolutionary. Whistle-blower. Mechanic. Crook. These are just a few of the people who had central roles in those events.

Behind every twist and turn in science is a person — perhaps brilliant, selfless and inspirational, or fickle, ambitious and egotistical. Each has their own story to tell. Collectively, those stories are woven into the fascinating fabric of scientific research that this publication probes and reports. So in this issue, our last in 2011, *Nature* has chosen to tell the stories of ten people who made a major difference to our — and, we hope, your — world this year (see page 437).

They have varied tales. Some demonstrate the sheer excitement of discovery: John Rogers, whose work is making electronics into wearable accessories, and Dario Autiero, whose team's claim that neutrinos can travel faster than light will be remembered for its glorious stretching of the imagination, even if the result doesn't ultimately hold up. And although the existence of the Higgs boson hangs in statistical limbo, the sheer buzz of its (near) discovery is enough for us to recognize Mike Lamont — the engineer who, perhaps more than anyone else, has kept particles whizzing around the Large Hadron Collider, and data churning out.

The role of hero is taken by Essam Sharaf, the engineer who temporarily took charge of the government in Egypt, whereas the villain is Diederik Stapel, a psychologist who perpetrated scientific fraud on a breathtaking scale — and in doing so underlined the difficulty of identifying wrongdoing in research. To represent those who stood by

science, we chose Lisa Jackson, whose efforts to promote evidence-based environmental regulation as head of the US Environmental Protection Agency have met with hostility in the nation's Republican-led Congress.

There are challengers: Tatsuhiko Kodama, who damned the Japanese government's studies of the radioactive fallout from the Fukushima nuclear disaster, and Rosie Redfield, who is using her blog to document her own attempts to replicate contentious claims about 'arsenic-based life'. And then there are those who are facing challenges of the future: Sara Seager, who, in a year punctuated by discoveries of distant exoplanets, is designing instruments to identify Earth-like worlds closer to our Solar System; and Danica May Camacho, one of the babies chosen to represent the world's population reaching a staggering 7 billion, and to raise awareness of the challenges to survival and sustainability that this milestone poses.

We can't pretend to have identified the only science newsmakers of this year, or even the most important. *Nature*, after all, is staffed by people with passions and foibles, and the selections reflect their subjective take on events. How did we decide on the final ten? We asked for suggestions from editors who deal with research manuscripts, as well as reporters and editors who put together *Nature's* News and Comment sections. We made long lists and short lists. We made them again. We argued in meetings. We discarded some obvious candidates and replaced them with figures whose stories had not been so widely told. We made a last-minute substitution two days before the section went to press.

Whatever you think of our ten, we hope that their stories provoke, remind, inspire and entertain. We also welcome feedback: we invite readers to nominate their own newsmakers of the year, and to vote in our online poll (go.nature.com/1w1xtk). We hope to repeat this exercise in years to come, and are already looking forward to the characters that we — and you — will meet in 2012. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhnuq



Particle physics is at a turning point

The discovery of the Higgs boson will complete the standard model — but it could also point the way to a deeper understanding, says Gordon Kane.

Let's tentatively agree that the Large Hadron Collider (LHC) detectors ATLAS and the Compact Muon Solenoid have discovered a Higgs boson, with a mass of about 125 gigaelectronvolts (GeV). Although standard statistical measures might not consider the situation settled, it seems very likely that there has indeed been a discovery at CERN, Europe's high-energy physics lab near Geneva, Switzerland, given that two quite different detectors both see a signal of some significance at about the same mass, and that both see the expected signals in two or more channels.

This is a profound turning point in the quest for a fundamental unified theory of the physical world. The properties and mass of the LHC's Higgs boson suggest that physicists will soon find superpartners for particles, and that we have begun to connect string theory to the real world.

The Higgs boson, an as-yet-unknown kind of matter thought to generate mass in other particles, is the final ingredient needed to complete and confirm the standard model of particle physics. This amazing theory describes the particles (quarks and leptons) and the strong, weak and electromagnetic forces that interact to make our world (with the addition of the theory of gravity). Quarks combine to make protons and neutrons; protons and neutrons to make nuclei; nuclei and electrons (a type of lepton) to make atoms, then molecules and chocolate and people and planets and stars and so on. The standard model has no puzzles or problems, and incorporates at a fundamental level everything from condensed-matter physics to astrophysics. It achieves the goals of four centuries of physics. The Higgs itself has been sought for decades: the main route through which its signal was reported at the LHC was the particle's decay into two photons. Collaborators and I first studied this signal in the mid 1980s, as a possible method for detecting the Higgs boson at the Superconducting Super Collider, which was to be built at Waxahachie in Texas but was cancelled in 1993.

Besides completing the standard model, the discovery of the Higgs tells us that a future, deeper underlying theory of the law(s) of nature must include and account for fundamental Higgs bosons. (Physicists have suggested alternative theories that include oddities such as composite Higgs bosons, but the CERN discovery essentially excludes them.) That will extend the standard model, and go beyond it to illuminate issues such as supersymmetry and the origin of dark matter.

A major and unexpected clue to the future offered by the CERN discovery is that the reported Higgs boson signal seems to behave as if it were a 'standard-model Higgs boson'. Under the standard model, this should not be possible, because relativistic quantum field theory shows that the Higgs' mass

must experience quantum corrections that are much, much larger than the mass itself. Because the masses of quarks, leptons and the W and Z bosons that mediate the weak force are themselves dependent on the Higgs mass, the standard model predicts masses for them many orders of magnitude larger than what we observe.

This can be fixed. When the standard model is extended to a supersymmetric theory, the nature of the predicted Higgs boson changes. Its mathematical behaviour improves and the resulting theory is realistic.

Physicists thought that a Higgs boson, when discovered, would take this supersymmetric form, so how have we discovered one so apparently identical to the impossible standard-model version? Working out how to interpret this could be a large step towards the underlying broader theory that will extend the standard model.

One explanation could come from an unexpected source: string theory or its extension, M-theory. Contrary to what you may have heard, predictions about the real world can be made from string theory, although the 10- or 11-dimensional theory must first be 'compactified' to 4 dimensions (with 6 or 7 small dimensions left curled up). There has been considerable progress on that, as well as on how to stabilize the fields that describe the curled-up dimensions.

My collaborators and I have shown that in generic string and M-theories — consistent with constraints from cosmology and incorporating the Higgs mechanism for generating mass — the lightest Higgs boson behaves very much like the standard-model Higgs boson. And it has a mass of about 125 GeV, just as observed.

We first reported these results at the international String Phenomenology Conference in Madison, Wisconsin, in August; and just days before the CERN data were reported, we posted a paper containing a significantly more precise prediction (G. Kane *et al.* <http://arXiv.org/abs/1112.1059>; 2011).

The same string theory (actually M-theory) that predicts the Higgs mass correctly also predicts that a spectrum of superpartners and some of their associated signals should now be discovered at the LHC. Particles such as gluinos — superpartners to gluons, which mediate the strong force — have not yet been searched for explicitly in the decay modes predicted by the string theories, mainly decay to top and bottom quarks. They could be found in these modes by the middle of next year. If so, the discovery may have a lower profile than the news of the Higgs boson, but the implications could be even greater. String theory could have come of age at last. ■

Gordon Kane is director emeritus of the Michigan Center for Theoretical Physics in Ann Arbor, Michigan.
e-mail: gkane@umich.edu

THE PROPERTIES AND
MASS OF THE
HIGGS BOSON
STRONGLY SUGGEST
THAT WE HAVE BEGUN
TO CONNECT STRING
THEORY TO THE
REAL WORLD.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/dyxd9g

SEVEN DAYS

The news in brief

POLICY

Patient protection

Rules to protect people taking part in federally funded research in the United States are adequate but should be made stronger, according to a 15 December report from a panel advising US President Barack Obama on bioethical matters. The Presidential Commission for the Study of Bioethical Issues recommended greater public access to data about studies on human subjects, and called for a system to compensate people who sustain research-related injuries. The report had been requested by Obama in 2010, after the discovery that US government-funded scientists had intentionally infected subjects with syphilis in a study in Guatemala in the 1940s. See go.nature.com/jzycz for more.

Cold shutdown

The three reactors at Japan's Fukushima nuclear plant that had meltdowns in early March have now been brought to a state of 'cold shutdown', the government announced on 16 December. The term means that the coolant in the reactors is stable and below its boiling point — and is usually an indication that any immediate crisis is over. But because the reactors are leaking, the milestone means very little in practice: water will still need to be pumped into the reactors to cool their decaying fuel, and residents who once lived near the plant will not be able to return until the land has been decontaminated. See go.nature.com/cdoruh for more.

Forest suspense

Final voting on a law that would relax forest protection rules in the Brazilian Amazon was last week delayed until March 2012, giving

environmentalists who oppose the changes more time to make their case to the Brazilian Congress. Scientists fear that the altered 'forest code' would weaken rules on tree-clearing that have reduced deforestation in the Amazon. The bill has already been passed by Brazil's Senate. See go.nature.com/bdfot2 for more.

Chimp research

Most biomedical research on chimpanzees is "unnecessary", the US Institute of Medicine found in a report released on 15 December. The report means that research using chimps that is funded by the US government will be sharply curtailed. The report

was immediately accepted by the US National Institutes of Health, where director Francis Collins said that "something like 50%" of the agency's roughly 37 projects involving chimps would be phased out because they do not meet the report's criteria. See page 424 for more.

NASA science head

John Grunsfeld, an astrophysicist and astronaut who carried out repairs on the Hubble Space Telescope, will head NASA's US\$5-billion science mission directorate, the agency confirmed on 19 December. *Nature* reported the first news of the appointment in November. Starting on 4 January 2012, he

replaces Ed Weiler, who retired in September. See go.nature.com/6pg9vv for more.

Australian reshuffle

Australian Prime Minister Julia Gillard took the nation's science leaders by surprise last week when she demoted her science minister, Kim Carr, to a non-cabinet position. Carr, who was well liked by the science community, had held the innovation, industry, science and research portfolio since 2007. Those areas were split in Gillard's reshuffle: science and research were added to the tertiary-education portfolio under Senator Chris Evans, and climate-change minister Greg Combet was given



CORNELL UNIV

Cornell wins New York science campus

Look out Boston and San Francisco. New York City has signed off on a US\$2-billion science and engineering campus to stimulate high-tech industry and inspire new start-up companies. On 19 December, New York's mayor, Michael Bloomberg, chose a bid by Cornell University and the Technion-Israel Institute of Technology in Haifa, which he called the "boldest and most

ambitious" of seven applications received in a year-long contest. With \$450 million in hand, a new campus (artist's impression pictured) is slated to be built on Roosevelt Island by 2017. It will admit 75 full-time faculty members and 300 graduate students; an expansion to 280 faculty and 2,000 students is planned over the next three decades. See go.nature.com/733z3e for more.

F. KRAUS responsibility for innovation and industry. See go.nature.com/6jaak5 for more.

RESEARCH

From air to orbit

Rockets bound for orbit could one day be fired not from launch pads, but from the underbelly of the largest aeroplane ever built — according to Stratolaunch Systems, a company that was announced on 13 December. Stratolaunch, which is headquartered in Huntsville, Alabama, has influential founders: aerospace pioneer Burt Rutan (who won the Ansari X Prize in 2004 for designing the *SpaceShipOne* rocket-plane, the first private craft to reach suborbital flight) and Microsoft co-founder and billionaire Paul Allen. They say that the air-launch-to-orbit system (intended to carry both cargo and human payloads) would be cheaper and more flexible than the conventional launch-pad approach. See go.nature.com/ead2ql for more.

Tiniest frogs

This tiny adult female frog (*Paedophryne dekoti*, pictured) is the world's smallest tetrapod, according to Fred Kraus at the Bishop Museum in Honolulu, Hawaii, who published his finding on 12 December (F. Kraus *ZooKeys* **154**, 71–88;



2011). Just 8.5–9.0 millimetres long from snout to vent — about a millimetre shorter than other tiny frog species — the amphibian was found living in leaf litter in forests in Papua New Guinea. A large number of diminutive frogs live in the region, which Kraus says may be a biological oddity. However, he also points out that miniature frogs are hard to find in the field, so their presence elsewhere could have been overlooked.

A twin Earth

NASA's Kepler telescope has reached one of its major mission milestones: discovering another Earth-sized planet. It has also spotted a Venus-sized planet in the same star system, which is about 290 parsecs (946 light years) away from us, researchers reported on 20 December at a press conference and in *Nature* (F. Fressin *et al.* *Nature* <http://dx.doi.org/10.1038/nature10780>; 2011). Although

scientists were delighted with this feat of detection, both planets orbit far too close to their parent star to be habitable. See go.nature.com/i372nd for more.

Variome project

A project to log all the genetic variations that cause disease in humans took a step forward last week with the launch of its Chinese arm at a meeting in Beijing. The Human Variome Project, based in Victoria, Australia, hopes to collect genetic data from laboratories all over the world, and share it in international databases of genes and diseases. China is taking on about a quarter of the project's estimated 20,000 genes, and will set up a genetics institute in Beijing to coordinate activities and give training in genetic counselling and testing. See go.nature.com/intbvt for more.

BUSINESS

Preventing HIV

Pharmaceutical firm Gilead Sciences wants to sell anti-HIV drugs to healthy people, to reduce their chances of becoming infected with the virus. On 15 December, the company, based in Foster City, California, filed an application to the US Food and Drug Administration to sell its two-in-one antiretroviral medication Truvada

COMING UP

31 DEC–1 JAN

NASA's twin GRAIL spacecraft are due to ease into orbit around the Moon, from where they will start to map lunar gravity in March 2012.

go.nature.com/msewft

1 JANUARY

Regulation of aviation's greenhouse-gas emissions is set to start in the European Union. All flights to or from Europe will have to buy carbon credits under the region's emissions trading scheme. China and the United States are contesting the plan.

(emtricitabine/tenofovir disoproxil fumarate) to people not infected with HIV. Such 'pre-exposure prophylaxis' (PrEP) has been supported by clinical trials (see *Nature* **476**, 260–261; 2011). The US Centers for Disease Control and Prevention in Atlanta, Georgia, is currently working on guidelines for administering PrEP and monitoring patients who are taking the therapy.

Amgen shake-up

Two top executives at the biotechnology giant Amgen are leaving after more than a decade in charge, the company said on 15 December. Chief executive Kevin Sharer and head of research and development Roger Perlmutter are both retiring, replaced respectively by chief operating officer Robert Bradway and chief medical officer Sean Harper. Amgen, based in Thousand Oaks, California, posted US\$15.1 billion in sales last year; its products include biological drugs for rheumatoid arthritis and anaemia.

➔ NATURE.COM

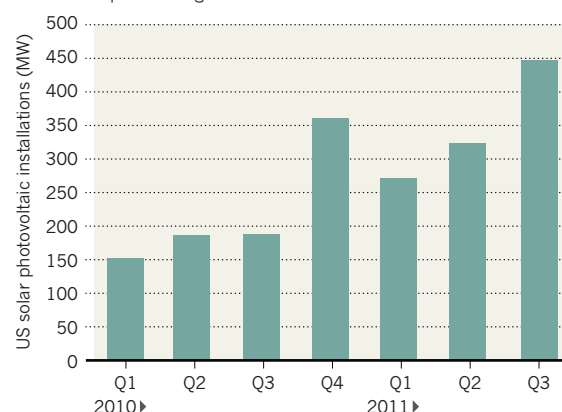
For daily news updates see: www.nature.com/news

TREND WATCH

US solar-energy installations spiked during the third quarter of 2011, according to a joint report by the Solar Energy Industries Association in Washington DC and GTM Research, headquartered in Boston, Massachusetts. Declining module prices have aided the boom, but the rush was also driven by a federal programme that provides solar-investment tax credits in the form of cash grants. Although the tax credit will continue until 2016, the cash-grant programme is due to expire at the end of this month.

US SOLAR SPIKE

Falling prices for photovoltaic modules helped the United States to a record quarter for grid-connected solar installations.



NEWS IN FOCUS

US BUDGET Funding bill backs translational-medicine centre for the NIH **p.423**

ETHICS Report sets a high bar for NIH research involving chimps **p.424**

YEAR IN REVIEW From the Arab Spring to unruly neutrinos **p.426**



NEWSMAKERS Meet 10 key people behind the events of 2011 **p.437**

P. LATRON/INSERM



Some researchers argue that work on a new strain of avian flu should be conducted only in labs with the highest biosafety precautions (above).

BIO SAFETY

Fears grow over lab-bred flu

Scientists call for stricter biosafety measures for dangerous avian-influenza variants.

BY DECLAN BUTLER

It is a nightmare scenario: a human pandemic caused by the accidental release of a man-made form of the lethal avian influenza virus H5N1.

Yet the risk is all too real. Since September, news has been circulating about two groups of scientists who have reportedly created mutant H5N1 variants that can be transmitted between ferrets merely breathing the same air, generally an indicator that the virus could also spread easily among humans.

The work raises the spectre of a disease that spreads as fast as ordinary seasonal flu, but with a fatality rate akin to wild-type H5N1 — an order of magnitude greater than the

mortality rate of roughly 2.5% seen during the catastrophic flu pandemic of 1918.

Until now, debate about the new variants has focused on whether the research poses too great a security risk to be published — even if partially redacted — a question currently under consideration by the US National Science Advisory Board for Biosecurity (NSABB).

A number of scientists argue, however, that the NSABB's deliberations have come far too late. Because further research on the new variants now seems inevitable,

➔ NATURE.COM
Read more at
Nature's avian flu
special:
go.nature.com/pkzzbu

a far more important question, they say, is whether the labs that hold samples of the virus — and those who will seek to work with

them in the future — have sufficient biosafety protection to make sure it cannot escape.

"This horse is out of the barn," says Richard Ebricht, a molecular biologist and biodefence expert at Rutgers University in Piscataway, New Jersey. "At this point, it is utterly futile to be discussing restricting the publication of this information," he adds, pointing out that the results have already been seen by many flu scientists, including referees, and are probably spreading through the flu grapevine faster than a speeding neutrino.

Sources say that one of the studies, led by Ron Fouchier of Erasmus Medical Center in Rotterdam, the Netherlands, has been submitted to *Science*, and that the other, led by Yoshihiro Kawaoka of the University ▶

SAFETY BY DEGREES

The precautions needed when working on a pathogen are determined by the threat it poses.

Biosafety level (BSL)	Typical precautions	Examples
1: Agents not associated with consistently causing disease in healthy adults	Lab coat and gloves	<i>Bacillus subtilis</i>
2: Agents causing mild, treatable human disease	Biosafety cabinets; limited lab access	Rabies, polio, seasonal flu
3: Agents causing serious human disease with possible treatment: high individual risk, low community risk	Decontamination of waste and lab clothing	Tuberculosis, anthrax, SARS
3 'Enhanced': Borderline BSL-3/BSL-4	Lab in isolation zone; dedicated power and air systems	Reconstructed 1918 flu virus
4: Agents likely to cause serious human disease with no treatment: high individual risk, high community risk	Full-body, air-supplied, positive pressure suits	Ebola, Nipah virus

► of Wisconsin, Madison, has been sent to *Nature*. (*Nature's* journalists do not have access to submitted manuscripts or the journal's confidential deliberations on them.) Fouchier also presented his results in September at the annual European Scientific Working Group on Influenza conference in Malta.

The mutant strains were not born out of a reckless desire to push the boundaries of high-risk science, but to gain a better understanding of the potential for avian H5N1 to mutate into a form that can spread easily in humans through coughing or sneezing. Some virologists have suggested that any genetic changes that made it more transmissible would probably blunt its deadliness. The new work seems to contradict that comforting idea. The studies should also help boost surveillance for similar changes in wild-type strains, and to develop diagnostics, drugs and vaccines.

Both experiments were conducted in labs rated at 'biosafety level 3 (BSL-3) enhanced' (see 'Safety by degrees'). Such labs require scientists to shower and change clothes when leaving the lab, and include other safety features such as negative air pressure and passing exhaust air through high-efficiency particulate

air filters. This should be quite sufficient to provide protection against an accidental release of the virus, some virologists say.

"Current biosafety rules are adequate for safely doing such transmission experiments with H5N1 viruses or any other influenza virus," says Peter Palese, a virologist at Mount Sinai School of Medicine in New York.

Requiring the more stringent protocols of BSL-4 facilities would hamper the research needed to develop countermeasures against an H5N1 pandemic, says Masato Tashiro, a virologist at the National Institute of Infectious Diseases in Tokyo, because it would limit the number of researchers able to work with the virus. As such, he believes that the work should be done in BSL-3 enhanced facilities.

HIGH SECURITY

But others say that to protect not only the researchers working on the viruses, but also society at large, the new H5N1 variants must be restricted to BSL-4 labs. These labs have far tougher safety and security measures, such as requiring workers to wear positive air pressure suits and undergo more rigorous decontamination; some also have additional security measures, such as video surveillance and bomb-proofing. Corraling this research in BSL-4 facilities would also immediately limit the proliferation of the viruses in labs, because only a few dozen such facilities exist worldwide, says Ebright. Indeed, one regulatory official, who requested anonymity, says that he is most concerned about the H5N1 mutants being handled in BSL-3 labs in countries with weak biosafety cultures or competences.

Deborah Middleton, an H5N1 researcher at the high-containment facilities at the Australian Animal Health Laboratory in Geelong, says that the characteristics of the new variants "fulfil the criteria of a BSL-4 pathogen", adding that she believes they would probably be handled as such in her institution. Indeed, the original experiments to create the viruses should also have been conducted in a BSL-4 facility, argues Hervé Raoul, director of the Jean Merieux-INSERM BSL-4 lab in Lyons, France.

Past experience suggests that the risk of the new variant H5N1 escaping from a lab

is far from negligible. Over the past decade, severe acute respiratory syndrome (SARS) has accidentally infected staff at four high-containment labs in mainland China, Taiwan and Singapore, variously rated as BSL-3 and BSL-4. A US National Research Council report released in September detailed 395 biosafety breaches during work with select agents in the United States between 2003 and 2009 — including seven laboratory-acquired infections — that risked accidental release of dangerous pathogens from high-containment labs.

And the rapid spread of an escaped flu virus would make it more dangerous than other deadly pathogens. "When SARS or BSL-4 agents get out, their potential for transmission on a global basis is quite limited," says Michael Osterholm, who heads the University of Minnesota's Center for Infectious Disease Research and Policy in Minneapolis, and is a member of the NSABB. "Influenza presents a very difficult challenge because if it ever were to escape, it is one that would quickly go round the world."

Fouchier declined to comment on these biosafety issues, saying only that his experiments had been reviewed by authorities in the Netherlands and the United States where "H5N1 virus is a class-3 agent because antivirals and vaccines are available". Kawaoka did not respond to interview requests.

Some scientists say that they are looking to the World Health Organization (WHO) to provide timely leadership in this biosafety debate. But Gregory Hartl, a spokesman for the WHO in Geneva, Switzerland, says the agency is unable to comment, because it has not yet seen the written studies. Meanwhile, the NSABB has not said when it will publish its advice. In a statement to *Nature*, the US Department of Agriculture said that it (and the US Department of Health and Human Services) will conduct any appropriate technical review of the new H5N1 variants.

Ebright laments that important questions of biosafety and biosecurity are largely left to the discretion of individual researchers. "In the United States, there is only voluntary oversight for biosafety, and with the exception of the select agents rule, there is no oversight of biosecurity," he says. Given the choice, says Middleton, flu researchers often resist working in higher biocontainment levels simply because they would no longer have the convenience of doing their research in BSL-3 labs at their own institutes, and because working in a BSL-4 lab is inherently more difficult.

The situation contrasts sharply with the barrage of legislation to regulate research that involves placing human subjects at risk, notes Ebright, where proposed projects are rigorously reviewed before they can start. "What's remarkable," says Ebright, is that for research of this type on H5N1, "which puts at risk not one individual but potentially hundreds, thousands or millions of individuals, there is no oversight whatsoever". ■

MORE ONLINE

TOP STORY



● India makes plans to phase out classroom dissections go.nature.com/dlpenn

MORE NEWS

- Young researchers lose out in EU funding programme go.nature.com/liqcoy
- Statistical method finds hidden patterns in complex data go.nature.com/ggflri
- How bacteria break a magnet go.nature.com/hyweco



NIH director Francis Collins seems to have secured government backing for a translational-science centre.

FUNDING

Last-minute wins for US science

Bill tops up health, energy and translational-science spending.

BY IVAN SEMENIUK AND SUSAN YOUNG

A year-end push by the US Congress to pass a \$915-billion spending bill has delivered on a key initiative proposed by Francis Collins, director of the National Institutes of Health (NIH).

In addition to funding the NIH at nearly \$30.7 billion for the 2012 fiscal year, the bill dissolves the National Center for Research Resources (NCRR) and allocates a large portion of its budget — roughly 45% — to the creation of a National Center for Advancing Translational Sciences (NCATS). It is the biggest organizational change at the biomedical agency in decades.

The translational-science centre was proposed by Collins late last year and fast-tracked by the administration of President Barack Obama for establishment in 2012. But the dissolution of the NCRR has met with some resistance within the NIH, particularly from those worried about the fate of its programmes, which will be parcelled out to other NIH institutes and the director's office (see *Nature* 471, 15–16; 2011). And in June, Denny Rehberg (Republican, Montana), chairman of the House Committee on Appropriations for Labor, Health and Human Services, Education, and Related Agencies objected to not having

received a formal budget request for the new centre. The committee made no provision for it in a draft bill released on 29 September. The reappearance of the centre in legislation passed by the Senate on 17 December — and expected to be signed into law by Obama — brings Collins's initiative closer to reality.

"Having the NIH in the discussions about the future of translational research is critical, so we are excited about its potential role," says Margaret Anderson, executive director of FasterCures, an organization based in Washington DC that works to speed up medical progress. "This is a smart allocation of resources."

The NIH also gets a reprieve from footing

the \$298-million cost of the Global Fund to Fight AIDS, Tuberculosis and Malaria, which will now be covered by the Department of State. Overall, the NIH receives \$299 million more than last year.

The omnibus bill also includes allocations for the Department of Energy and the Environmental Protection Agency (see 'Science by numbers'). Funding for some agencies, including NASA and the National Science Foundation (NSF), was established in a bill that passed on 17 November (see *Nature* 479, 455–456; 2011).

The Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, will also see a small increase from last year's allocation, some of which comes from the mammoth health-care reform bill that Obama championed in 2010. But Karl Moeller, executive director of the Campaign for Public Health in Washington DC, worries about a longer-term decline in the agency's discretionary budget and about the CDC's lack of champions on Capitol Hill. "If you poll the public, they love prevention, but funding that, for some reason, just doesn't excite members of Congress," says Moeller.

Energy research will see a small rise in 2012. The bill allocates \$4.9 billion to the Department of Energy's Office of Science — roughly 1% over its 2011 figure — and \$275 million to the Advanced Research Projects Agency-Energy (ARPA-E) to pursue high-risk, high-return projects. The latter figure is far below the president's ambitious request of \$550 million for ARPA-E, but it will allow the fledgling agency to proceed with funding a new round of projects (see *Nature* 471, 145–146; 2011).

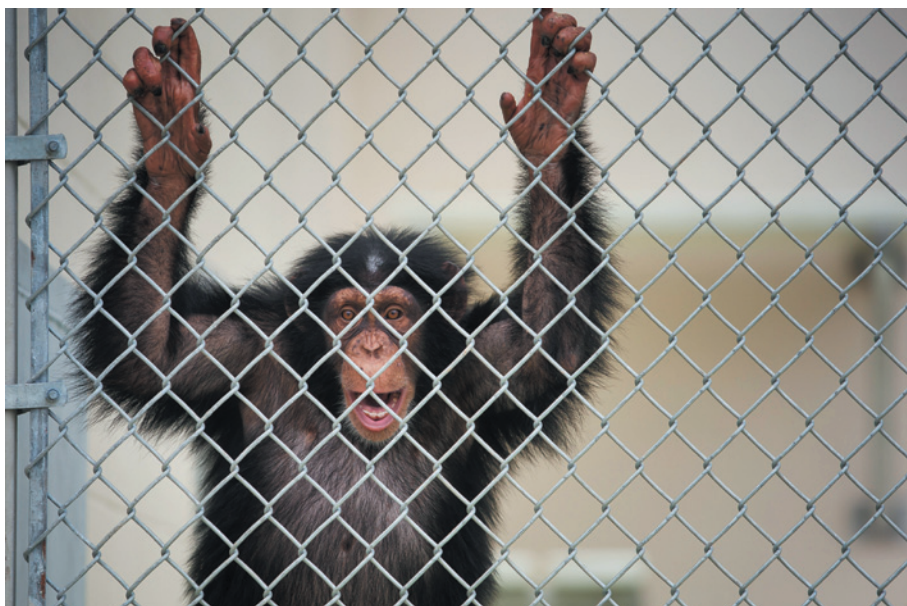
Climate science has come under particular scrutiny, a reflection of Republican antipathy towards the regulation of greenhouse-gas emissions. The bill requires the administration to provide a detailed account of money spent on programmes related to climate change in 2011. The Environmental Protection Agency's budget has been cut by 3.4%, with a \$14-million reduction in funding for clean-air and climate-research programmes. The bill does, however, make \$10 million available to the state department to contribute to the Inter-governmental Panel on Climate Change and the United Nations Framework Convention on Climate Change. ■

SCIENCE BY NUMBERS

Selected figures from the latest spending bill passed by the US Congress, compared with the president's 2012 request and previous fiscal year allocations.

	2010 actual	2011 estimated	2012 requested	2012 passed
National Institutes of Health	30,784	30,399	31,456	30,698
Environmental Protection Agency	10,281	8,682	8,973	8,463
Centers for Disease Control and Prevention (discretionary)	6,390	5,649	5,817	5,668
Department of Energy's Office of Science	4,964	4,858	5,416	4,889
US Geological Survey	1,112	1,084	1,118	1,070

Sources: White House OMB, House Committee on Appropriations, CDC



Tightened standards could pose a barrier to all but a few studies making use of research chimps.

ANIMAL EXPERIMENTATION

Chimp research under scrutiny

US-government-funded studies to be evaluated under stricter criteria.

BY MEREDITH WADMAN

A landmark report published by the Institute of Medicine (IOM) last week has triggered the US government to reconsider what, if any, chimpanzee research should receive funding. It has also left Christopher Walker pondering the implications for an enterprise close to his heart — developing a preventive vaccine against the hepatitis C virus (HCV).

Walker, director of the Nationwide Children's Hospital's Center for Vaccines and Immunity in Columbus, Ohio, found himself agreeing with many of the recommendations in the report, which was released on 15 December. They include strict criteria for assessing when chimps can be used in research projects (see 'Raising the bar for chimp research'), and the establishment of an expert committee to judge proposed studies on a case-by-case basis.

Walker even agreed that no chimp experiment aimed at developing therapies for the 170 million people worldwide infected with HCV would meet those criteria. For example, although the chimp is the only other animal susceptible to HCV infection, new *in vitro*

culture systems and small-animal models containing human cells offer alternatives.

What gave Walker pause was an issue that split the otherwise-unanimous IOM authors. Only half felt that chimpanzees may be necessary to identify HCV vaccine candidates for use in humans. The split makes it hard to predict whether government-funded HCV vaccine researchers will still be approved to work with chimps. "If you have an idea for an entirely new approach to vaccination, but can't get proof of that principle in animals, my fear is that it will never move ahead into human trials," says Walker, who is studying protective immunity to HCV in chimps.

Although the IOM report leaves the door

partly open for HCV vaccine research, in other disease areas it is less equivocal. Jeffrey Kahn, the committee chair and deputy director of policy and administration at Johns Hopkins University's Berman Institute of Bioethics in Baltimore, Maryland, notes that only one other active research area meets the IOM standards: existing studies in which chimps are used to produce monoclonal antibodies and to test their safety. "No other current biomedical research areas were seen as meeting the committee's criteria," Kahn says.

During a media briefing after the report's release, Francis Collins, the director of the National Institutes of Health (NIH) in Bethesda, Maryland, which supports 612 chimps and oversees most of the government's chimp research, said that 37 agency-funded chimp studies would be evaluated under the report's criteria. Those that do not pass muster will be phased out; no new studies involving experimentation on chimps will be funded unless the criteria are met.

DIVIDED OPINION

The committee's split over the use of chimps to identify promising HCV vaccine candidates reveals a fault line within the biomedical community. In a letter sent to the committee on 25 September, 100 scientists, including several Nobel laureates, argued that, without access to the chimp, there will be a "substantial delay" in creating an effective vaccine, and that there is no guarantee that promising-looking alternatives, such as mouse models, will prove to be effective.

Yet the report's conclusions reflect the tenor of the times, as well as the progress of alternatives to animal research. Public opposition to the use of chimps in research has grown steadily. This is reflected in proposed legislation on Capitol Hill that would ban invasive and comparative-genomic chimp research, and in a petition from animal-welfare groups being considered by the US Fish and Wildlife Service to declare research chimps as endangered, which would effectively end any chimp research not aimed at conservation.

The IOM report was requested by the NIH after high-profile politicians and activists protested against the agency's plan to move 186 semi-retired chimps back into active research (*Nature* **467**, 507–508; 2010). The committee wrote that its deliberations, although formally considering only scientific necessity, were "suffused with an awareness

RAISING THE BAR FOR CHIMP RESEARCH

Experiments will need to comply with these criteria to win funding from the National Institutes of Health.

Biomedical studies

- No other suitable model is available
- The study cannot be performed ethically in humans
- Important advances would be significantly slowed or prevented if the study was not done
- Chimps must be housed in appropriate physical and social environments or in natural habitats

Comparative-genomics and behavioural studies

- The study provides otherwise unattainable insight
- The study is performed on acquiescent animals using minimally invasive techniques in a manner that minimizes pain and distress
- Chimps must be housed in appropriate physical and social environments or in natural habitats

of the moral cost of such research”.

Researchers are divided over the number of experiments that stand to be affected. Collins guesses that “something like 50%” of the 37 NIH-funded experiments would be curtailed, but the directors of the three NIH-supported research centres that conduct invasive experiments say that they carried out only seven such experiments with NIH funded chimps this year.

The paucity of invasive experiments, they say, shows that chimps are already being used highly selectively. “There are some assumptions, it seems, that the IOM criteria were not being applied already. And I think that’s incorrect,” says Thomas Rowell, director of the New Iberia Research Center in Louisiana, the largest US chimp research centre.

Two of four projects using chimps at the Michale E. Keeling Center for Comparative Medicine and Research near Bastrop, Texas, this year involved testing treatments on chimps chronically infected with hepatitis C. A third studied hepatitis-C infectivity and the fourth involved HIV/AIDS. All four are either complete or near completion, so are unlikely to be affected by the new recommendations.

John VandeBerg, director of the Southwest National Primate Research Center in San Antonio, Texas, says that there are two NIH-funded studies under way there. “It is my judgement that both of these studies would qualify for continuation” under the IOM criteria, he says.

But animal-welfare activists are pleased that the committee came down resoundingly on the side of less is better. Even in the case of a preventive HCV vaccine, says Kathleen Conlee, senior director for animal-research issues at the Humane Society of the United States in Washington DC, “it didn’t say, ‘We absolutely need chimps for this’, but rather, ‘We couldn’t decide’”.

Supporters of chimp research take heart from a separate IOM observation that a new or re-emergent disease could require the use of chimps in the future. “The IOM report is the first authoritative document to acknowledge the very real possibility that new and emerging threats to human health may require access to chimps,” says Christian Abee, director of the Keeling Center. “A plan will have to be developed to ensure that chimps are available in the future.”

When asked whether the new report will require a reassessment of the chimp-breeding ban that the NIH put in place in 1995, Collins says that he does not have an analysis showing how many chimps should “be kept on hand should the need arise from some sort of new pandemic for their use”. But, he says, “with more than 600 chimps already available and owned by the NIH, it seems as though we have a pretty substantial population to work with.” ■

POLICY

Rules on integrity signal tighter stance

Many US agencies now have policies banning political manipulation of research — but ambiguities remain.

BY EUGENIE SAMUEL REICH

When Kathleen Sebelius, secretary of the US Department of Health and Human Services, overruled the Food and Drug Administration (FDA) on 7 December and denied over-the-counter access to the Plan B One-Step ‘morning-after’ pill for girls aged under 17, advocates for scientific integrity in government were outraged. Although the FDA’s decision to allow access had been based on an in-depth scientific review, Sebelius — who is not a scientist by training — claimed that the data did not support the view that young girls would be able to use the drug safely. “The key problem is she re-reviewed the science,” says Francesca Grifo of the Union of Concerned Scientists in Washington DC.

It is no small irony that a row over political interference in scientific decision-making should erupt just as President Barack Obama’s initiative to shield government scientists from such pressures comes to fruition. Agencies and departments across the US government have been working to submit final drafts of scientific-integrity policies, many of which make some reference to disallowing politically motivated alteration of data.

The policies were due to be submitted to the government’s Office of Science and Technology Policy (OSTP), where its director John Holdren is overseeing the process, by 17 December. Last week, *Nature* conducted its own survey of the effort by contacting agencies directly. Of the 11 departments and agencies that have confirmed to *Nature* that they are drafting new policies, six now have public policies that make some reference to forbidding politically motivated alteration of data. A seventh, the Department of Justice, has told *Nature* that a working draft does so.

But three agencies — the Department of Health and Human Services, the Department of Energy and the Department of Defense — have not made their policies public or answered *Nature*’s questions about them. A fourth agency — the National Institute of Standards and Technology (NIST) — has no public policy but has told *Nature* that a working draft does not explicitly ban political alteration of data. Henry Wixon, the chief counsel for NIST, says that this has not traditionally been an issue at the agency,

but adds that he may consider amending the NIST draft policy. Similar shifts have taken place at NASA and the National Science Foundation, both of which included language banning political interference in the latest versions of their integrity policies, released in the past week. Earlier drafts did not include such language.

The stakes are highest at regulatory agencies where science directly informs policy. Among these, the Environmental Protection Agency has moved forward with asserting that the work of its scientists and engineers should be free from political influence. However, critics have complained that the agency’s draft policy could be clearer about the fact that the requirement applies to agency political appointees, not just to scientists employed by it. Another regulatory agency, the energy department, received a letter on 15 December from the Project on Government Oversight, a watchdog group based in Washington DC, calling for its policy to include an independent oversight mechanism for breaches of research ethics.

Yet some question whether the integrity policies go far enough. It is far from clear, for example, whether such a policy at the Department of Health and Human Services would have prevented the furore over Plan B.

Nick Steneck, a research ethicist at the University of Michigan in Ann Arbor, says that the OSTP should offer stronger leadership to make it clear that integrity policies must address political interference. “Issuing a simple document is not sufficient,” he says. ■ [SEE EDITORIAL P.413](#)

CORRECTIONS

The graph of a hypothetical stock portfolio in the News story ‘Blockbuster drug bows out’ (*Nature* **480**, 16–17; 2011) omitted the effects of dividends and stock splits; instead of losing 39%, it should have risen by 82%. Similarly, US\$100 worth of Pfizer stock in 1986 would have grown to \$2,139 by 1999 and shrunk to \$1,629 by the end of 2010.

The News Feature ‘Master of illusion’ (*Nature* **480**, 168–170; 2011) wrongly attributed the idea “if there’s something you can be certain of in this world, it’s that your hand is your hand” to Descartes instead of G. E. Moore.

365 DAYS:

2011 *in review*

BY RICHARD VAN NOORDEN

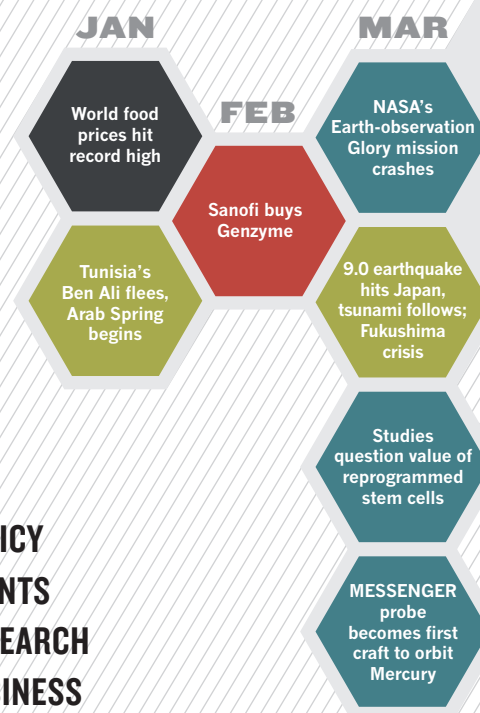
For science — as for politics and economics — 2011 was a year of upheaval, the effects of which will reverberate for decades. The United States lost three venerable symbols of its scientific might: the space-shuttle programme, the Tevatron particle collider and blockbuster profits from the world's best-selling drug all came to an end. But the year also saw stirrings of science's future: hopes that research might blossom following the Arab Spring; cheap vaccines rolling out in Africa; and the first fruits of genome sequencing being used in the clinic. All this was overshadowed by the triple trauma of Japan's devastating earthquake, tsunami and nuclear meltdown, and a continual chipping away at science funding as nations struggled with the financial crisis.

THE STEM-CELL ROLLER COASTER It has been an emotional year for advocates and opponents of embryonic stem (ES) cell research. In July, there was relief when a federal judge dismissed a lawsuit seeking to halt US government funding of research using human ES cells. But in October, the European Court of Justice banned patents based on human ES cells, although the ruling's impact on European science remains unclear. And November brought a shock when the first company to test a human ES-cell product in patients — Geron of Menlo Park, California — bailed out of the field. Back in the lab, ES cells in a dish were coaxed into forming complex three-dimensional tissues, including a retina and a pituitary gland. Scientists also managed to create a human ES-cell line by using cloning technology on a human egg — the only catch being that the cells contained an extra set of chromosomes. Meanwhile, scientists' early love affair with induced pluripotent stem (iPS) cells gave way to a more nuanced assessment of their abilities. In the first half of the year, a string of worrying research papers showed that the reprogrammed adult cells can trigger adverse immune reactions in mice and may contain genetic abnormalities. But other studies demonstrated potential benefits: iPS cells derived from patients can be used to study their diseases in a dish, with published examples including a rare condition linked to accelerated ageing and neurological disorders such as schizophrenia.

R. HAHN/FERMI LAB



Powered down: the Tevatron collider was turned off for good in September.



THE YEAR IN BRIEF

PUZZLING OVER PARTICLES What a cliff-hanger. After gathering evidence from around 420 trillion proton-proton collisions in the Large Hadron Collider at CERN, Europe's particle-physics lab near Geneva, physicists announced in December that they still couldn't confidently state whether or not the Higgs boson exists — although they did identify its most likely hiding place, at a mass of around 125 gigaelectronvolts. That mystery joined another messy plot thread: the astonishing claim that neutrinos can travel faster than light. In September, a team at the OPERA experiment — a neutrino detector at the Gran Sasso National Laboratory near L'Aquila, Italy — claimed to see a beam of neutrinos make the 730-kilometre journey from CERN 60 nanoseconds quicker than light could have managed, apparently breaking the cosmic speed limit. Although physicists have tried to poke holes in the result, no one has so far spotted an obvious flaw, leaving a world on tenterhooks as independent experiments attempt to replicate the finding. The mysteries continued with dark matter, as a flurry of results from astrophysical experiments provided a welter of contradictory and confusing signals. But at least one story came to a definite conclusion this year: after more than 25 years spent colliding particles, the massive Tevatron accelerator at Fermilab in Batavia, Illinois, was switched off for good.

BATTERED REPUTATIONS No year is free of scientific scandal, but 2011 saw some particularly high-profile cases. In October, prominent psychologist Diederik Stapel was fired by Tilburg University in the Netherlands after an investigation committee found extensive fraud in his work (see page 441). Three months earlier, evolutionary psychologist Marc Hauser resigned from Harvard University following last year's misconduct findings, but while the US Office of Research Integrity continues to investigate the case, exact details of what he did remain remarkably scant. The scandal involving cancer geneticist Anil Potti, who resigned from Duke University in Durham, North Carolina, in 2010, reached new



NATURE.COM
For an interactive
round-up of the
year's events see:
go.nature.com/2011

heights in September, when patients enrolled in clinical trials based on his science brought a lawsuit against the university and the scientists involved, claiming that they had been harmed. And November saw the dramatic arrest and brief jailing of Judy Mikovits, prominent for her work (now partially retracted) linking chronic fatigue syndrome to a virus.

ARAB AWAKENING Elated scientists joined jubilant revellers throughout Egypt on 11 February, when Hosni Mubarak resigned after 30 years as the nation's president. He stepped down a few weeks after popular uprisings forced out Tunisian President Zine al-Abidine Ben Ali. With dictators ousted, many researchers remain optimistic that the Arab Spring's new freedoms will lead to progress in science, education and democracy. But by the time Libya's ruler, Muammar Gaddafi, was killed in October, it was increasingly clear that change would be slow in coming — and would depend heavily on those who take power. One immediate effect of the revolutions was to throw archaeology into turmoil: foreign archaeologists had to leave both Libya and Egypt, and Egypt's Supreme Council of Antiquities was left paralysed after its controversial but charismatic leader, Zahi Hawass, was forced to leave office in July.

A DRYING DRUG PIPELINE The world's best-selling drug, Lipitor (atorvastatin), racked up more than US\$100 billion in sales for its maker Pfizer over the past 14 years. But that golden age ended when the patent expired on 30 November, and with more blockbuster drugs soon to lose their patent protection, a familiar question loomed: why is research producing fewer profitable drugs to take their place? All the usual coping

strategies were in evidence this year: companies slashed research budgets, and forged external collaborations with academia and with other firms. Most notably, in February, Sanofi struck a \$20-billion deal to acquire the pioneering biotechnology firm Genzyme of Cambridge, Massachusetts. But some innovative treatments brightened the picture: boceprevir and telaprevir were the first two in an anticipated wave of hepatitis C drugs to win approval. They were joined by belimumab, the first drug approved to treat lupus in half a century, and ipilimumab, a treatment for advanced melanoma, which extended the small family of cancer drugs targeted at patients with a particular genetic profile. A fundamental breakthrough came from researchers led by Brian Kobilka at Stanford University in California. They captured the crystal structure of a key player on the surface of cells — the G-protein-coupled receptor — locked in an embrace with its protein partner. Between one-third and one-half of all drugs target these receptors, so understanding how they work could give industrial research and development a much-needed boost.

SCIENTISTS UNDER PRESSURE In a year in which some scientific societies came under fire for shifting their human-rights work away from traditional campaigning on

behalf of oppressed researchers, there were many examples of scientists who needed that support. In Iran, for instance, physics student Omid Kokabee was jailed and charged with spying; and in France, French-Algerian physicist Adlène Hicheur began his third year in jail without having faced trial. In the United States, the administration of President Barack Obama did make progress on a much-vaunted effort to draft

2,326

**NUMBER OF CANDIDATE
EXTRASOLAR PLANETS
IDENTIFIED BY NASA'S KEPLER
SPACE TELESCOPE; 207 ARE
ABOUT THE SIZE OF EARTH.**



The end of the space-shuttle programme: *Atlantis* landed on 21 July.

policies on scientific integrity, a response to charges of censorship under his predecessor George W. Bush. The policies would give scientists at US government agencies more freedom to discuss their work (see page 425).

BRAVE NEW WORLDS After 30 years and 135 missions, the space-shuttle era is over. *Atlantis* touched down for the final time on 21 July, leaving NASA reliant on Russian Soyuz capsules to get astronauts off the planet. The space agency also struggled to contain the spiralling costs — \$8.7 billion and rising — of the James Webb Space Telescope, whose funding was narrowly saved amid budget cuts. Although NASA's Glory probe, to measure solar output, crashed on take-off in March, other robotic craft soared. The Curiosity rover is en route to Mars, and the Juno probe is heading for Jupiter. The MESSENGER spacecraft entered orbit around Mercury, and the Dawn mission did likewise around Vesta in the asteroid belt. Closer to home, Aquarius began to measure the saltiness of Earth's oceans from orbit. Meanwhile, China

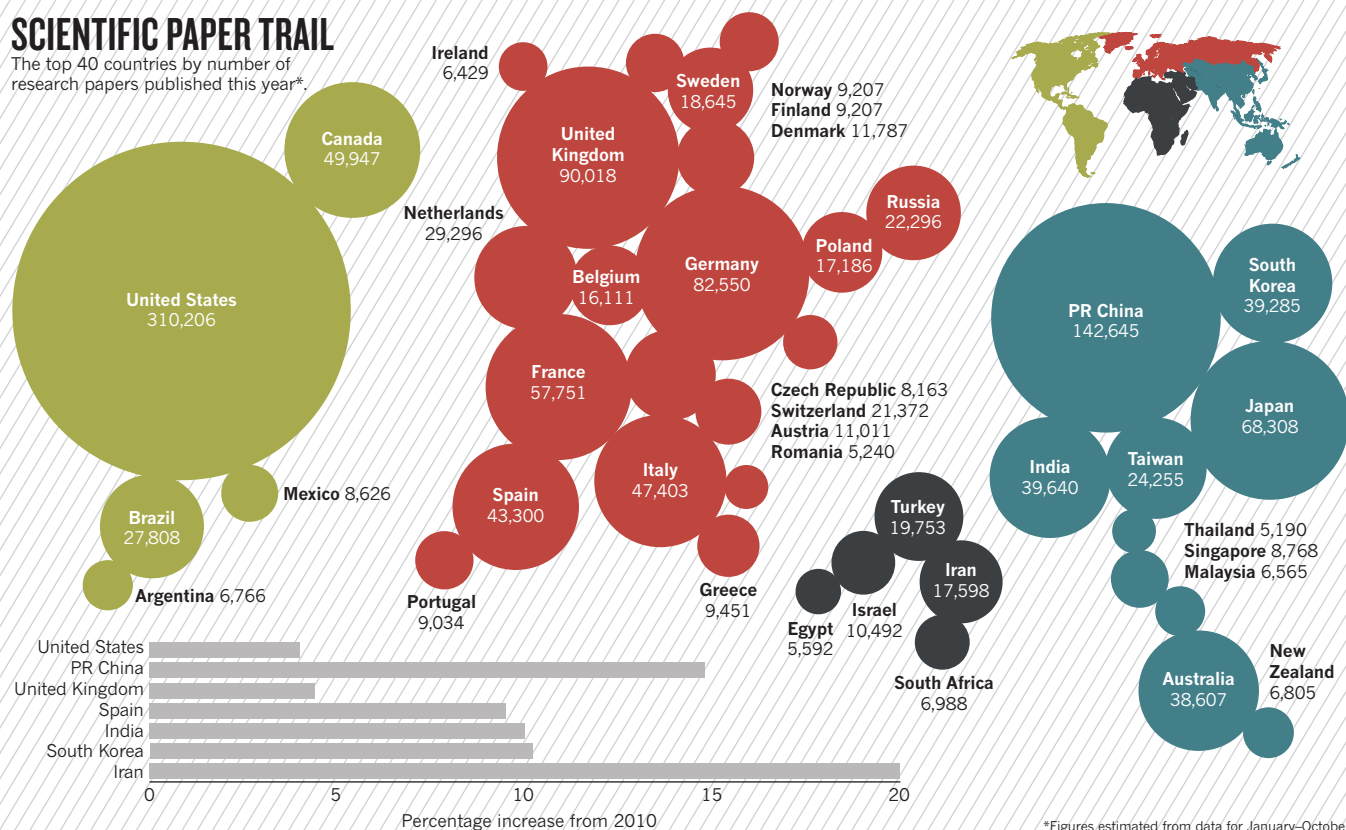
bolstered its growing power in space, achieving its first space docking, and Russia tasted failure with its botched Mars mission, Phobos-Grunt. Europe finally lofted the first satellites of its Galileo navigation system into orbit. But the most thrilling space research was the flood of data on new extrasolar planets emerging from NASA's Kepler mission and from ground-based telescopes. So far, more than 700 new worlds have been identified — although none is quite Earth's twin.

SEQUENCING: CHEAP AS CHIPS Care for a fish-and-chip genome? The cod and the potato were among the many organisms unravelled by sequencing machines this year, but the most fascinating genomes were those that revealed humanity's past. The first complete genome of an Aboriginal Australian offered clues that pioneers left Africa for the south more than 60,000 years ago, and the genome of the Black Death pathogen *Yersinia pestis*, recovered from victims' teeth, shed light on how the disease laid waste to Europe in the fourteenth century. Bacterial outbreaks nowadays can be sequenced as they happen — as shown by the rapid identification of the *Escherichia coli* strain that wreaked havoc in northern Europe in the spring. That feat was achieved in three days by one of a new wave of chip-based sequencers, which manufacturers say will provide genomes for less than \$1,000. Yet for all their promise, some makers of small and large sequencing devices announced miserable profits in the year's third quarter, after biologists, fearful of falling research funding, held off buying new equipment. Instead, sequencing companies are eyeing up new markets in biomedicine, promoting the use of genetic tests and whole-genome sequencing to diagnose clinical diseases, which is starting to have life-changing effects.

CUTS, CUTS, CUTS We've had the stimulus: now prepare for austerity. As governments pondered where and when to slash budgets, scientists once again rallied to make the case for funding basic research. After the United States narrowly avoided a government shutdown in April, most of its federal science agencies saw modest funding increases

SCIENTIFIC PAPER TRAIL

The top 40 countries by number of research papers published this year*.





Fallout: inspectors take stock of the damage to Fukushima Daiichi.

3.5×10^{16} bq

AMOUNT OF THE RADIOACTIVE ISOTOPE
CAESIUM-137 RELEASED BY THE FUKUSHIMA
DAIICHI POWER PLANT — ABOUT HALF THAT
RELEASED FROM CHERNOBYL.

GI. VERLINI/AEA

proposed for 2012 — but the prospect of across-the-board cuts looms over 2013. In the United Kingdom, cuts to facilities made in 2010 began to bite as researchers lost funds for buildings and major equipment, although some of the cash was dribbled back in concessions through the year. In Spain, the Prince Felipe Research Centre, a flagship biomedical-research facility in Valencia, was forced into severe cutbacks, and the San Raffaele Scientific Institute in Milan, Italy, is accepting a rescue package backed by the Vatican after running up a debt of US\$1.5 billion. Non-governmental organizations also suffered: the Global Fund to Fight AIDS, Tuberculosis and Malaria said that it would not fund new grants until 2014 because of budget cutbacks. The dizzying growth of China's economy, however, meant that its science budget expanded by 12.5% this year. And the European Commission was clearly feeling optimistic when it requested €80 billion (US\$104 billion) for research under the Horizon 2020 proposals unveiled in November.

IN THE SHADOW OF FUKUSHIMA Even Japan, the nation best prepared for a tsunami, was overwhelmed by the monster waves that struck the coast of Sendai on 11 March, following a magnitude-9.0 earthquake. Tens of thousands of people died, and hundreds of thousands were displaced. But it was the meltdown of three tsunami-damaged reactors at the Fukushima Daiichi nuclear plant — the worst nuclear disaster since Chernobyl — that soon gripped the world's attention. Fortunately, wind carried much of the radioactivity out to sea. It took nine months before the reactors could be declared safely in cold shutdown; and it will take decades and hundreds of billions of dollars to clean up the plant. Unsurprisingly, energy policy in Japan — and in Germany, Italy and Switzerland — has turned abruptly against nuclear power. So where will the world find its low-carbon energy? Many countries are banking on recently discovered reserves of natural gas locked up in shale, accessible thanks to 'fracking', in which rock is fractured with high-pressure fluids. But concerns that the process pollutes air and water sparked vociferous protests in the United States and a ban in France. Solar power went through turmoil too, as slow demand for solar panels, oversupply of products and materials, and widespread subsidy cuts bit deep into the industry's profits. Several companies

declared bankruptcy — including high-profile US start-up firm Solyndra of Fremont, California. The good news for consumers, however, is that solar-power modules are rapidly getting cheaper.

LIVING IN THE ANTHROPOCENE The world's population passed 7 billion this year, and our carbon emissions carried on rising too. Little wonder that geologists gathered in May to discuss whether human impact on the planet deserved recognition through the declaration of a new geological epoch: the Anthropocene. The year's main climate-change conference in Durban, South Africa, ended with a last-minute agreement to work towards a global treaty to set emissions-reduction targets, bolstering political progress that included Australia instituting a carbon tax and a record low for deforestation in the Brazilian Amazon. For the 3.4 billion people surviving on less than \$2 a day, getting enough food and avoiding disease are more urgent priorities. And it is not getting any easier: after droughts and oil-price hikes, the cost of food staples reached a record high in February and dropped only slightly through the year. Help from the GAVI Alliance, however, came in the form of a new pneumococcal vaccine rolled out widely in February, and a rotavirus-vaccination programme to target childhood diarrhoea in Africa. In October, interim results from a huge phase III clinical trial of the RTS,S malaria vaccine candidate showed disappointing efficacy and no effect on mortality. But there was promising news in HIV: one trial showed that treating infected people with antiretroviral drugs early helped to prevent them spreading the virus, and two others suggested that giving antiretrovirals to healthy people reduced their chances of becoming infected. ■

IMAGES OF THE YEAR

Science saw farther than ever this year, glimpsing new wonders in the microworld and in the distant reaches of the Solar System. But some of the most powerful images were taken at an all too human scale, as the earth shook, volcanoes blew and mankind continued to threaten other species' survival. Here are 2011's most striking pictures.

Images selected by *Nature's* art and design team
Text by Daniel Cressey



365 DAYS:
the year in science

JAPAN'S PAIN

The shattered village of Ōtsuchi testifies to the savagery of the magnitude-9 earthquake and subsequent tsunami that hit Japan's Tohoku region on 11 March. It killed tens of thousands, left hundreds of thousands homeless and triggered a meltdown at the Fukushima nuclear plant.



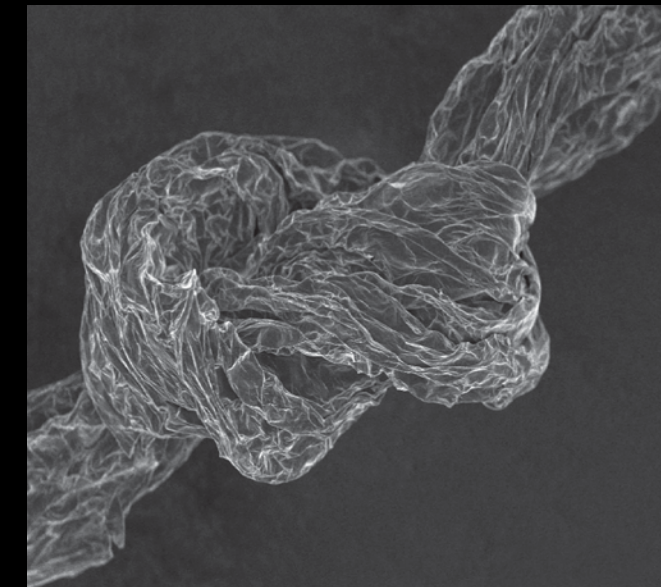
CAGED FURY

Rats don't deserve their bad name, but this ball of fury won't win over many muropobes. Russian scientists bred this aggressive rat strain to compare it with more docile creatures in a study on domestication that has teased out several genetic regions linked to tame traits.



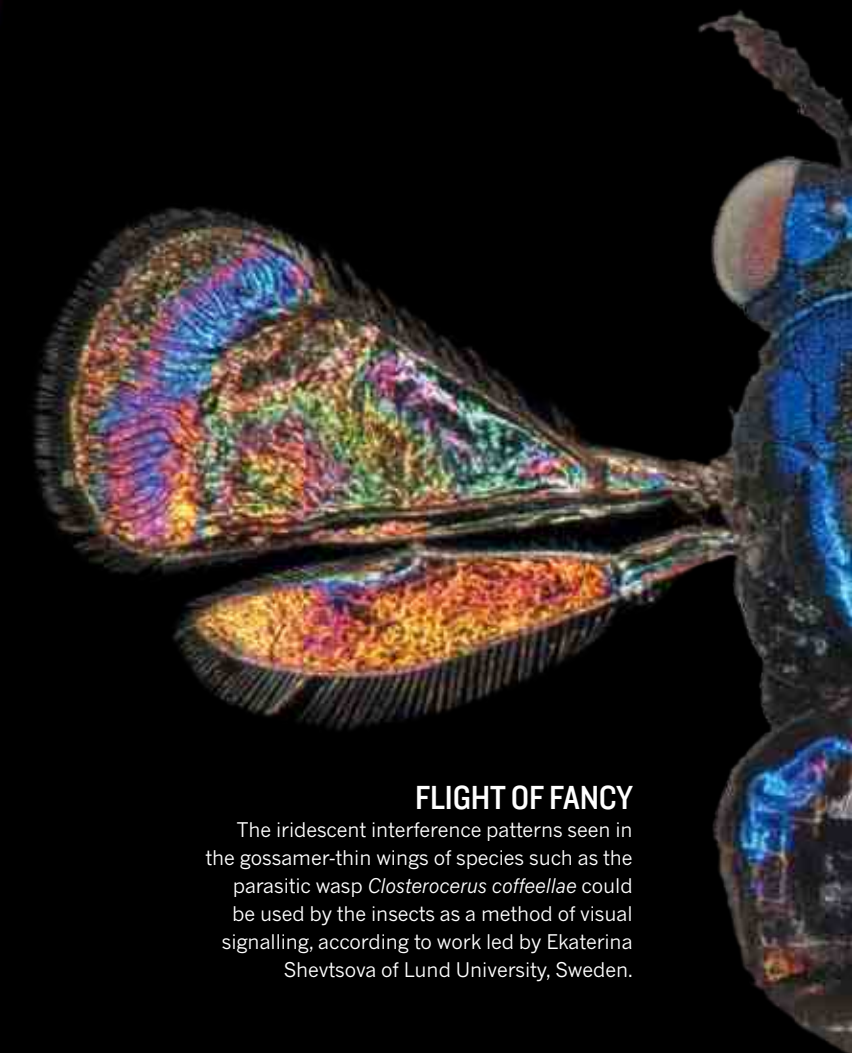
CELLULAR CHRISTMAS

Donna Stolz created a festive wreath by assembling images of mammalian cells from more than 25 experiments. The picture adorned her Christmas card last year and won recognition for the University of Pittsburgh biologist in the 2011 Nikon Small World photography contest.



UNFORGETTABLE

This 400- μm -long knot in graphene, tied by Zhen Xu and Chao Gao at Zhejiang University in China, shows exquisite control at the nanoscale. Xu and Gao spun flat liquid crystals of graphene oxide into flexible fibres metres long, and then converted them into graphene threads.



FLIGHT OF FANCY

The iridescent interference patterns seen in the gossamer-thin wings of species such as the parasitic wasp *Closterocerus coffeellae* could be used by the insects as a method of visual signalling, according to work led by Ekaterina Shevtsova of Lund University, Sweden.

FINAL ENDEAVOUR

We shall never see this sight again. Thirty years, 135 flights and two disasters after the space shuttle first launched, *Endeavour* — pictured here in May docked to the International Space Station — and the rest of the shuttle fleet are retiring to museums.



DOWN TO EARTH

With the demise of the shuttles, Russia's venerable Soyuz capsule is for now astronauts' sole route to the space station. Looking more like a nineteenth-century diving bell than a twenty-first-century spacecraft, this capsule brought three astronauts down to Kazakhstan in November.



DAWN PATROL

Vesta's scarred face was captured by NASA's Dawn craft as it orbited the 530-kilometre-wide denizen of the asteroid belt. The origin of the crater at the centre of this image, which contains a mountain 15 kilometres high, is one of many mysteries Dawn will probe during its year-long reconnaissance.

ULTRA-FLYWEIGHT

Balanced on a dandelion, this ordered metal lattice could usurp other, structurally disordered, ultra-light materials. Tobias Schaedler of HRL Laboratories in Malibu, California, and his team created it by plating nickel-phosphorus onto a polymer microlattice, and then removing the polymer skeleton.



RHINO AIR

A vertiginous helicopter ride took this black rhino on the first leg of a journey from its shrinking habitat in eastern South Africa north to Limpopo. But its cousin the western black rhino was less fortunate: it was one of 13 species and subspecies officially declared extinct this year.



BOOM!

The beauty of a plume from the Puyehue–Cordón Caulle volcano complex in Chile was lost on the thousands of people forced to flee the area, and on those in neighbouring Argentina who were showered in ash from the eruption in June.

ISS: P. NESPOLI/NASA; SOYUZ: B. INGALLS/NASA; VESTA: NASA/JPL-CALTECH/UCLA/MPS/DLR/IDA; MICROLATTICE: D. LITTLE/HRL LABORATORIES; RHINO: WWF/GREEN RENAISSANCE; VOLCANO: C. SANTANA/AFP/GETTY; RAT: V.J. MUSE/NATIONAL GEOGRAPHIC/STOCK JAPAN; D. SAGOL/REUTERS/CELLS; D. STOJUNIV. PITTSBURGH/NIKON SMALL WORLD; GRAPHENE: C. GAG; WINGS: E. SHEVTSOVA



CARL DETORRES

Ten people who mattered this year.

365 DAYS:
the year in science

DARIO
AUTIERO

RELATIVITY CHALLENGER

The shy experimentalist whose team claims to have found faster-than-light neutrinos is happy for the work to stand or fall.

BY NICOLA NOSENGO

Dario Autiero can hardly keep track of his e-mails any more: hundreds keep pouring in from the media and his fellow physicists across the globe. “But the real problem is science amateurs,” says Autiero, who works at the Institute of Nuclear Physics in Lyons, France. “They send e-mail upon e-mail saying that they had predicted it all.”

Autiero has been at the centre of this media storm, scientific scepticism and amateur theorizing since 23 September, when he and his colleagues at the international Oscillation Project with Emulsion-Tracking Apparatus (OPERA) experiment announced results that seemed to remove a cornerstone of modern physics. At a seminar at CERN, the particle-physics laboratory near Geneva, Switzerland, and in a paper posted on the arXiv.org website (<http://lanl.arxiv.org/abs/1109.4897>), the OPERA team described how neutrinos — fundamental particles with no electrical charge and very low mass — seemed to make the 730-kilometre journey from CERN to an underground laboratory at Gran Sasso in Italy some 60 nanoseconds faster than the speed of light would allow. If true, the result will challenge Albert Einstein’s theory of special relativity and force theoretical physicists to rewrite their textbooks.

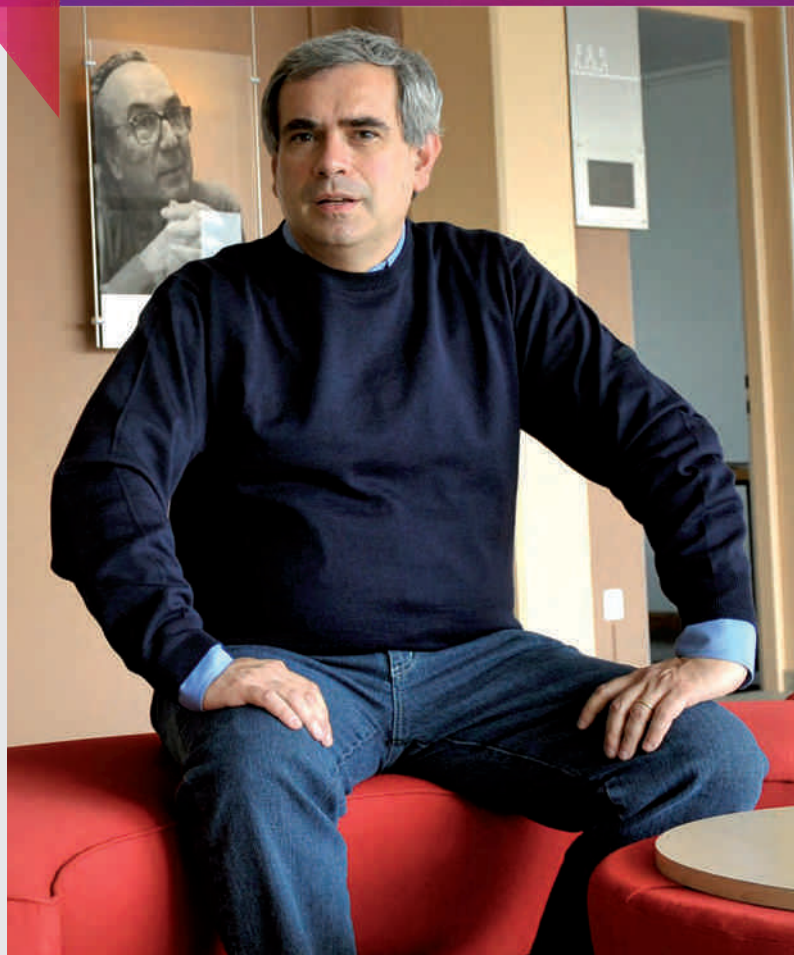
Autiero was caught off guard by the frenzy that ensued. “The media have a completely different timescale,” says Autiero, a quiet, shy man who is clearly uncomfortable in the spotlight. “They want answers right away, whereas as a scientist I am used to spending years looking for them.”

The experiment’s day job is to study how the three known types of neutrino ‘oscillate’, or switch between different identities, as they travel between Geneva and Gran Sasso. But Autiero realized that it also offered a chance to measure the particles’ speed with unprecedented precision. This would require complicated statistics to match the neutrinos detected in Italy to the proton collisions that generated them in Switzerland, as well as extreme accuracy in measuring the distance between the particles’ points of departure and arrival. In March, the team thought it had finally solved these problems — but that left them facing a much bigger one.

“When I first saw the data I was sure there was some mistake,” Autiero recalls. The team spent the following months reanalysing the statistics, recalibrating the instruments and having them checked by independent institutes. The 60-nanosecond discrepancy refused to go away. “At that point, we had done all we could do with the data,” Autiero says. “It was our duty to release them, so that others could look into them.”

The experiment’s advisory board felt the same. Still, a few OPERA members refused to sign the paper, and some scientists have criticized the team for telling reporters that the non-refereed preprint would be posted on arXiv.org. Astrophysicist Martin Rees of the University of Cambridge, UK, called the OPERA result an “embarrassment” and quoted the late astronomer Carl Sagan’s remark that “extraordinary claims require extraordinary evidence”.

But Autiero finds the criticism baffling. “If you think your experiment is done properly,” he says, “you should treat unexpected results as you would treat expected ones.” Even more puzzling to him was the flood



of papers about the experiment posted on arXiv in the following weeks. Some of them “made valuable critiques and forced us to refine our work”, he says. “But mostly they pointed at trivial problems we had solved months ago.” A hardline experimentalist who admits to being “religious” when it comes to the scientific method, he felt offended by some of those papers.

Autiero thinks that the matter will be settled by the end of 2012. His team is repeating the experiment with a variation in how the neutrinos are generated at CERN, to rule out that process as a possible source of errors. On 17 November, the group released an updated arXiv preprint that included an analysis of 20 neutrinos from the repeated experiment, and simultaneously submitted the paper for publication in the *Journal of High Energy Physics*. This new version confirmed the initial result, and won back many of the collaborators who had refused to sign the initial paper.

But the true test, Autiero says, will come from two neutrino experiments that are attempting to replicate the results: the Main Injector Neutrino Oscillation Search, in which neutrinos fired from Fermilab near Chicago, Illinois, travel 720 kilometres to a detector in a mine in Minnesota; and T2K, which sends a beam from Tokai on the east coast of Japan to a detector in Kamioka, 300 kilometres to the north. Even if they prove the OPERA result wrong, Autiero says that he would not consider the experiment a failure. “Whatever happens, our most important contribution is not the result — it’s the way we did the measuring,” he insists. “We have refined a method others can now use, and this is what science is about.” ■

M. TREZZINI/EPA/CORBIS

PLANET SEEKER

In a year of exoplanet excitement, one astronomer is already asking what comes next.

BY ERIC HAND

Hundreds of years from now, says Sara Seager, “people will look back at us, and they won’t remember me or you. They’ll remember us as the generation of people who first found the Earth-like worlds” outside our Solar System.

This year, scientists have come tantalizingly close, says Seager, an astronomer at the Massachusetts Institute of Technology (MIT) in Cambridge. NASA’s Kepler space telescope, which monitors thousands of stars for dimming caused by an orbiting planet, has found 28 confirmed exoplanets this year, including one that is slightly bigger than Earth and in its star’s habitable zone. More than 2,000 await verification.

But Seager, a member of the Kepler science team, wants to do better. Kepler can reveal a planet’s size and orbital radius, she explains. But to find out whether such a planet is Earth-like — with free oxygen or other signs of biological activity in its atmosphere — astronomers need a spectrum of the parent star’s light reflected from or transmitted through the atmosphere. Because the stars in Kepler’s field of view are up to 920 parsecs (3,000 light years) away, they are too dim for that.

Seager wants to search for Earth-like planets no more

than 30 parsecs away, close enough that their atmospheres could be studied. Her tool would be a $10 \times 10 \times 30$ -centimetre space telescope designed to watch a single star for a planetary transit. Such an ‘ExoplanetSat’ would not be able to analyse spectra by itself. For that, Seager will need an orbiting telescope such as the Terrestrial Planet Finder, an ambitious concept that NASA put on ice in 2006. But a fleet of ExoplanetSats could provide a resurrected planet finder with a map of where to look. Each ExoplanetSat would cost less than US\$1 million. Rather than a telescope mirror, it would rely on a modified, \$1,300 commercial lens. And dozens could be launched very cheaply, piggybacking on rockets carrying other missions.

“I’m trying to do new things,” says Seager, who is teaching herself the necessary engineering and is aiming for a 2013 launch. Her group has received roughly \$3 million in funding from MIT, Draper Laboratory in Cambridge and elsewhere.

Geoff Marcy, an astronomer at the University of California, Berkeley, and one of the first exoplanet seekers, lauds Seager’s creativity in unfamiliar fields. “There are thousands of scientists working on exoplanets,” he says. “She’s looking for something different.” ■

J. KNIGHT



**SARA
SEAGER**



LISA JACKSON

POLLUTION COP

The top US environment official has faced relentless attacks on the country's pollution regulations.

BY JEFF TOLLEFSON

Lisa Jackson knew from the start of 2011 that she would face a rough year as head of the US Environmental Protection Agency (EPA). Republican conservatives had just swept the board at the 2010 congressional elections, and they took over the House of Representatives vowing to handcuff the EPA, which they viewed as a prime example of big government run amok. One hostile lawmaker advised Jackson to get her own parking space on Capitol Hill because Congress would be calling her in regularly to defend her agency and the environmental agenda of President Barack Obama. Jackson has testified before Congress 11 times this year.

C. DHARAPAK/AP

The Republican hostility has its roots in one of Jackson's first major decisions after taking the helm of the EPA in 2009. A chemical engineer turned public servant, Jackson issued a scientific assessment formally declaring that greenhouse gases pose a threat to human health and welfare. That ruling set the stage for the EPA to regulate greenhouse-gas emissions under the Clean Air Act, and Jackson immediately went to work. Looking back, she says the ruling represents a long-awaited triumph of science after years of delay and obfuscation about climate research under the previous administration. "Restoring that science I hope will be among the hallmarks of the Obama EPA," she says.

The ruling has led to a series of decisions that have provoked

SCIENCE REVOLUTIONARY

An engineer was catapulted from Tahrir Square to Egypt's parliament and fought to rebuild science.

BY MOHAMMED YAHIA

As academics joined the millions protesting in Egypt's streets this spring, the voice of one engineer soon began leading chants. Essam Sharaf was in the thick of demonstrations in January, and he became the first prime minister of a post-revolution cabinet in March — promoting science as a solution to the country's woes. But by November, he had resigned amid a second surge of popular protest.

The 59-year-old Sharaf was born in Egypt and earned degrees in engineering from Cairo University and Purdue University in West Lafayette, Indiana. By 2010, he was an academic engineer at Cairo University and a fierce critic of Egyptian President Hosni Mubarak's regime.

Sharaf's stance during the uprising made him popular with the young

revolutionaries. He was high on their list of candidates to lead the new transition government, along with Nobel laureate Ahmed Zewail, a chemist from the California Institute of Technology in Pasadena. When Sharaf was chosen, hundreds of thousands of revolutionaries gathered to greet him in Tahrir Square. "If I can't bring the change you want, then I will return to the lines with you," he told them.

Once in office, Sharaf said that science could solve many of Egypt's dire developmental problems, ranging from water security to energy. His cabinet began drafting plans to improve the education system and, in June, he approved a long-standing proposal by Zewail to build a university of basic and applied research. "Countries do not move forward except with scientific research," said Sharaf at the launch of the project, to be called the Zewail City for Science and Technology. "Without question, he has been a visionary engineer," says Kumares Sinha, a professor of civil engineering at Purdue University who once supervised Sharaf. "Over the years he has been a strong advocate for science and engineering in the Arab world as a way to advance and reform the society."

But that success in advocating for science did not carry over to politics. Critics accused Sharaf's cabinet of weakness, believing that all real power lay in the hands of the military junta. When this discontent erupted into violent protests in November, Sharaf and his cabinet resigned. Elections took place in November for a new parliament, which should be in place early next year, and a presidential election will be held in June. For now, the future of science in Egypt is as uncertain as that of the reinvented country itself. ■

Republican ire. The Obama administration has worked with automobile makers over the past three years to establish fuel-efficiency and greenhouse-gas-emissions standards, the president's signature environmental achievement thus far. The EPA has also begun rolling out requirements for major industrial facilities and, despite delays, is preparing to issue the first-ever greenhouse-gas standard for US power plants and refineries.

Jackson's EPA has also targeted smog-forming pollutants as well as mercury and other toxic chemicals from industrial facilities and power plants. In October, the agency announced plans to regulate the waste water generated by shale-gas development, which involves injecting water and chemicals at high pressure into gas-bearing rocks, an activity that many fear could pollute groundwater resources.

The Republican Congress has fought these efforts, arguing that pollution regulations cost jobs and resources at a time when the country is short of both. As of November, by Jackson's count, Republicans had brought more than 170 attacks on basic environmental protections up for vote this year, although none of their efforts to weaken existing regulations was successful.

Even Jackson's critics acknowledge that she holds up well under fire and has succeeded in pushing through some significant environmental regulations. "She's been very effective," says Jeff Holmstead, an attorney at Bracewell and Giuliani in Washington DC, who headed the EPA's air-quality programme under former president George W. Bush.

Environmentalists, however, were hoping for much more from the Obama administration. Already angry about the slow pace of greenhouse-gas regulations, they were incensed when Obama overruled Jackson and quashed the EPA's proposed standards for ozone pollution. But most realize that an EPA chief has limited power, says Frank O'Donnell, who heads the advocacy group Clean Air Watch in Washington DC. He adds, "Jackson herself has done as strong a job as anybody possibly could under the circumstances". ■



**DIEDERIK
STAPEL**

FALLEN STAR

A psychologist's spectacular fraud became an example in open investigation.

BY EWEN CALLAWAY

Dutch social psychologist Diederik Stapel had been called one of the 'bright thrusting young stars' of the field before his career imploded this autumn over fraudulent research. In prominent studies that explored prejudices and stereotypes, Stapel didn't just fudge data, he fabricated entire experiments — seemingly for much of his career, according to a preliminary report issued on 31 October by the three university committees investigating his work. They are still sifting through data from approximately 150 published papers to catalogue Stapel's misdeeds for a final report to be issued next year.

Stapel's case is unusual not only for its scale, but also for the speed and transparency of the investigation. Officials at Tilburg University in the Netherlands, where Stapel was dean of the School of Social and Behavioural Sciences, announced his suspension and their plans to investigate in early September, just days after students alerted them to irregularities in his published data. Pim Levelt, director emeritus of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, headed the Tilburg investigation and says that the committee acted rapidly and openly in the interests of Stapel's former students and collaborators. "There are co-authors who are losing most of their publications, and that is really a disaster when you are starting a science career," he says.

Levelt says that he drew lessons from other investigations, including the high-profile case of Marc Hauser, the former Harvard University psychologist who resigned this year after being found guilty of eight counts of research misconduct. The details of what Hauser did, however, are still unclear. "I was not very pleased by the Hauser case. It was a bit secretive, the final report was very short, and it took them three years," Levelt says.

But Daniele Fanelli at the University of Edinburgh, UK, who studies research misconduct, says that the committees investigating Stapel had much more freedom. Stapel quickly confessed and even identified papers that contained fabricated data. (He has since returned his PhD to the University of Amsterdam and reportedly sought mental-health care.) In cases in which accused scientists dispute the charges, Fanelli says, universities have to keep the investigations under wraps. ■

J. BUUS/PERSBUREAU VAN EIJNDHOVEN

K. HAMRA/AP

**ESSAM
SHARAF**



ROSIE
REDFIELDCRITICAL
ENQUIRER

A blogger's quest to replicate 'arsenic life' led to a remarkable experiment in open science.

BY ERIKA CHECK HAYDEN

She appeared like a shot out of the blogosphere: a wild-haired Canadian microbiologist with a propensity to say what was on her mind. And on 4 December 2010, what was on Rosie Redfield's mind was arsenic — specifically, a paper published two days earlier in *Science*, in which researchers funded by NASA claimed to have found bacteria that could incorporate arsenic into their DNA in place of phosphorus (F. Wolfe-Simon *et al.* *Science* **332**, 1163–1166; 2010). If true, the finding showed that life could be supported by a form of biochemistry radically different from the one we know. But Redfield's blog entry on the paper pulled no punches. "Basically, it doesn't present ANY convincing evidence that arsenic has been incorporated into DNA," she wrote.

Redfield kicked off a frenzy of criticism of the 'arsenic-life' paper in the blogosphere and the media. In June, *Science* published eight critiques of the paper. But that month, Redfield took matters into her own hands: she began attempting to replicate the work in her lab at the University of British Columbia in Vancouver, and documenting her progress on her blog (<http://rrresearch.fieldofscience.com>).

The result has been a fascinating story of open science unfolding over the year. Redfield's blog has become a virtual lab meeting, in which scientists from around the world help to troubleshoot her attempts to grow and study the GFAJ-1 bacteria — the strain isolated by Felisa Wolfe-Simon,

M. DEE/NATURE

CHILD OF
THE TIMES

The '7-billionth baby' was born into a world that is approaching a population plateau.

BY DECLAN BUTLER

On 31 October, the United Nations declared, the world's population reached 7 billion. Danica May Camacho, born on that day in the Philippines, was the first of several babies acclaimed for having nudged the population over the threshold. The choices were symbolic: the United Nations can at best only estimate that humanity will pass 7 billion sometime this year or next. But putting a face to the number drew attention to the challenges of absorbing a larger population and prompted a slew of Malthusian doomsayers to lament an overpopulated world.

Yet the underlying population story is not that there are now 7 billion people, nor that humanity's numbers will rise to somewhere around 10 billion by 2050. It is the dramatic slowing of population growth. The raw data reveal that the number of annual births, which had been

growing for centuries, peaked around 1990 at roughly 135 million, and has declined since then. "The world reached peak child before peak oil," says Hans Rosling, an epidemiologist at the Karolinska Institute in Stockholm. That is mostly because family size in the majority of poorer nations has been shrinking for decades, thanks to economic growth, improved family planning and decreased child mortality. Much of the developing world is closing in on the population-replacement fertility rate, about two children per woman.

Annual population growth has already dropped from 1.8% around 1950 to 1.1% in 2010 and is expected to reach zero around 2060–80. This has knock-on effects: in Africa, where the population has been growing fastest of late, the median age is set to increase, bringing most of the population into working age. This demographic dividend is likely to spur economic growth.

Even if the population problem is abating, the sheer number poses challenges for humanity and the planet. The remedies needed in the next decades are much the same as those needed today: reduce poverty to tackle the root cause of hunger and to accelerate the fall in population growth; develop sustainable agricultural practices that increase food production without gobbling up extra land, water and other resources; develop renewable energy sources and boost energy efficiency to deliver the power that the world will need while avoiding more global warming.

We may have defused the population bomb, but 1 billion people remain hungry and the planet's resources are stretched thin. ■

lead author of the *Science* paper and a microbiologist who worked in the lab of Ronald Oremland at the US Geological Survey in Menlo Park, California. For months, Redfield could not get the GFAJ-1 bacteria to grow reproducibly on a medium containing arsenic. Finally, in November, the bacteria took off. Redfield now plans to check whether they have incorporated arsenic into their DNA, but it will take even more work to show that they can survive without any phosphorus.

To Redfield, the exercise has shown how social media tools are binding science into a community closer than it has been since the early twentieth century, when it was possible for scientists to personally know everyone in their field. "Scientists are much more able to communicate with people we don't know, and to learn from people we've never met," she says.

Before arsenic life, Redfield was known in the evolutionary-microbiology community for championing the relatively unpopular idea that bacteria evolved the ability to take up DNA from their environment for the purposes of nutrition, not sex. So although much has been made of Redfield's criticism of Wolfe-Simon — a blogger on Gizmodo.com, for instance, photoshopped a picture of the two glaring lightning bolts at each other — Redfield says that her lonely defence of her own unpopular hypothesis has made her sympathetic to Wolfe-Simon. "I'm in that position; it's not that people doubt my experiments; they doubt my interpretation," she says. Like Redfield, many scientists now believe that the conclusions of Wolfe-Simon and her team were incorrect because the researchers didn't rule out the possibility that their cultures were contaminated with phosphorus. Wolfe-Simon's team has argued that any background phosphorus was insufficient to support the bacterial growth it observed.

Ford Doolittle, a biochemist who hired Redfield for her first faculty job and is now at Dalhousie University in Halifax, Canada, says that Redfield's work has proved a point by showing how science is supposed to work. "Science is way too uncritical of itself," says Doolittle. "We need more Rosies out there." ■



DANICA MAY
COMACHO



MIKE
LAMONT

THE HIGGS MECHANIC

The engineer who keeps the Large Hadron Collider running brought physics closer than ever to completing the particle zoo.

BY EUGENIE SAMUEL REICH

It was standing room only in the auditorium at CERN on 13 December, when physicists at the Large Hadron Collider (LHC) presented the best indications yet that the long-sought-after Higgs boson might have been found. But the man who keeps the particles colliding was absent. Mike Lamont heads the LHC operations group, which was holding its own workshop on the machine's function. The group of about 80 engineers staffs the LHC control room in shifts and coordinates hundreds of others working on the superconducting magnets, accelerating chambers and other equipment arrayed around the LHC's 27-kilometre ring near Geneva, Switzerland, all to keep the data flowing without overtaxing the machine.

This year, the LHC produced 500 trillion proton-proton collisions, 100 times more than in 2010, generating a torrent of data that has allowed scientists to collect suggestive — but not definitive — indications of a Higgs boson with a mass of about 125 gigaelectronvolts. Lamont had spent the year carefully edging up the number of collisions by injecting more protons into each of the bunches speeding around the ring, packing the bunches more tightly and tightening the colliding beams to an ever-smaller focus. He also had to solve day-to-day problems, such as radiation damage to beam-line electronics and 'UFOs' — unidentified falling objects — that occasionally plummet into the path of the beam and interrupt experiments.

Lamont, a physicist who joined the LHC preparation team in 2001, was among those who worked frantically to restart the machine after an accident in 2008 shut it down. He insists that getting the collider back online was all down to teamwork. His goal for the next year is to deliver three times more data than in 2011 — hopefully enough either to confirm that the hint of the Higgs was real, or to rule it out. "The experiments are like hungry chicks. They always want more," he says. ■



**TATSUHIKO
KODAMA**

FUKUSHIMA'S GADFLY

The emotional academic who challenged his government and took nuclear clean-up into his own hands.

BY DAVID CYRANOSKI

Tatsuhiro Kodama began his testimony calmly. But a few minutes into his speech before the Japanese parliament's health and welfare committee on 27 July, the biologist's tone grew sharp — and then downright angry — as he blasted the Japanese government for not accurately reporting the amount of radiation that had leaked from the Fukushima Daiichi nuclear power plant after the earthquake on 11 March. "This is clear negligence on the part of the government," he shouted. "With 70,000 people wandering around, unable to go home, what is the government doing?"

The 16-minute rant has since been viewed around one million times on YouTube, and Kodama, who is head of the Radioisotope Center of the University of Tokyo, quickly became known as the 'emotional scientist' spokesman for the victims of the Fukushima disaster. Journalists and local governments sought his advice on how to deal with the ongoing nuclear crisis, and he helped local governments to initiate some evacuation and decontamination efforts — all while the central government dawdled.

"He's been a real driving force," says Kaname Tajima, a ruling-party politician in charge of nuclear-disaster measures, who was stationed in Fukushima from June to September.

Sitting in his office in October, Kodama quietly recounts what led to his impassioned rant. He says that the information breakdown started

S. SHINDO

TECH EXEC

From flexible circuitry to miniature solar cells, this engineer has a knack for turning physics into technology.

BY ERIC HAND

The bulky crate in the back of the delivery truck didn't look like much, remembers John Rogers, thinking back to 1996 and his time as a postdoctoral fellow in physical chemistry at Harvard University in Cambridge, Massachusetts. But inside was his first commercial product, the InSite 300. It was designed to shine laser light on thin films and listen for an acoustic echo that would reveal their thickness and elastic properties without touching them. Active Impulse Systems, the start-up company he had co-founded a year earlier, had just sold its first InSite 300 for US\$400,000 to a California firm that made semiconductor tools. In that moment, as Rogers posed with his colleagues for a jubilant snapshot in the back of the truck, he knew

what he wanted to do with his life. "Physics for physics' sake is great," he says. "But if you can do physics for the sake of technology, I think that might be the way to go."

That attitude goes a long way towards explaining how Rogers came to win this year's Lemelson-MIT award for innovation — not to mention a MacArthur Foundation 'genius' fellowship in 2009. Now the head of a 40-person lab at the University of Illinois at Urbana-Champaign, Rogers still freely combines techniques from physics, chemistry, materials science and even bioengineering — and continues to pile up patents and spin-off companies. "That transition from a scientific idea to an engineering prototype — he's unbelievably good at it," says George Whitesides, who mentored Rogers during his postdoc.

A case in point is Rogers's work on flexible electronics, a crowded field that he joined about five years ago. The goal is to make electronic devices that can be worn rather than held — woven into clothes, say, or moulded to the body. But whereas many of his materials-science colleagues were working with organic materials, Rogers gave brittle silicon another chance. He found that ultra-thin silicon circuits printed onto an elastic surface could be highly flexible — and retain the benefits of silicon's low cost and high performance.

➔ NATURE.COM
Vote for your 2011
Newsmaker in our
online poll:
go.nature.com/ggvzox

Ones to watch

2012

DIETER EGLI

NEW YORK STEM CELL FOUNDATION

This year he reprogrammed a human egg and derived embryonic stem cells — but with three sets of chromosomes. Tweaks could produce ‘diploid’ lines.

PAUL MAHAFFY

NASA GODDARD SPACE FLIGHT CENTER

When the Mars Science Laboratory lands, the instrument that he runs will sniff for traces of methane, a gas that could be a clue to life.

ELON MUSK

SPACE X

His SpaceX Dragon capsule is scheduled to dock with the International Space Station early next year — a milestone for commercial space flight.

OSCAR PISTORIUS

SOUTH AFRICAN SPRINTER

His efforts to qualify for the 2012 Summer Olympics on carbon-fibre artificial limbs are set to reignite debate about technology’s impact on human performance.

ESKE WILLERSLEV

CENTER FOR GEOGENETICS, COPENHAGEN

This year, he tackled an Aboriginal Australian genome and extinct megafauna. Next year, his palaeogenetic sequencing ventures are bound to surprise.

in the first days after the disaster, when the government decided not to release data from an ¥11-billion (US\$141-million) system created to forecast how radiation spreads after a nuclear accident. The government claimed that because data were sparse, the system might cause unwarranted panic. That suggestion angers Kodama, who says that the Nuclear Safety Commission “should not be worried about confusion or panic. The specialist committee’s major mistake was trying to act as politicians rather than scientists.” Without the simulation results, which turned out to match later reports of a plume of radiation stretching to the northwest of the plant, evacuees from one high-radiation area fled to another.

Then, the Nuclear Safety Commission and the parliament bickered over whether safety levels should be set at 20 millisieverts or 1 millisievert, delaying decontamination efforts and further confusing citizens. “While these committees were arguing, the situation was getting worse and worse. That’s another thing that makes me mad,” says Kodama.

The academic eventually got the government’s ear. The week after his rant, Tajima visited him. The following week, Kodama met then prime minister Naoto Kan, and advised him to get more data from the worst-affected areas.

Kodama, who is responsible for the safe operation and maintenance of the University of Tokyo’s 27 isotope facilities, was already advising local officials about contamination. In late May, he started working in Minamisoma, a coastal city with around 70,000 inhabitants that straddles the

border of the 20-kilometre mandatory evacuation zone. On his counsel, the local government encouraged pregnant women and children, who face an increased risk from radiation exposure, to evacuate from those areas outside the exclusion zone that had elevated radiation. Later, such advisories became common in the wider affected region. Kodama also started emergency decontamination efforts in Minamisoma, teaching town administrators how to measure radiation and look for micro-hot-spots.

“You can’t just measure, you have to look for the source,” says Yoshiaki Yokota, a member of the local school board whom Kodama taught. “Before he showed me, I knew nothing about radiation.”

Kodama’s frustration continues. He says that the government is still not doing enough to help the victims, and he opposes plans to build a state-of-the-art ¥100-billion hospital in Fukushima city, arguing that support should be spread out more widely.

He also says that the government is still not releasing enough information. A ban on entry to the exclusion zone has kept scientists from sizing up the true situation in the area and has hampered the work of journalists. Kodama calls it a “censorship that is quite unusual in democratic countries”.

“The lesson from Chernobyl for the Japanese government is that there is so much psychological scarring, so we can never put too much information out there,” he adds. ■

One result of this work is a spin-off company based in Cambridge, Massachusetts, called mc10, that is working with the sporting-goods giant Reebok to roll out a product in 2012 that, Rogers says, will measure an athlete’s “kinetic health and well-being”. He wants to get inside the body, as well — mc10 is in the process of developing membranes studded with electrodes that can wrap around the brain or heart to provide neurologists and cardiologists with vastly improved diagnostic maps (J. Viventi *et al. Nature Neurosci.* **14**, 1599–1605; 2011).

Another spin-off co-founded by Rogers — Semprius, based in Durham, North Carolina — aims to make photovoltaic arrays that produce solar energy for less than 10¢ per kilowatt-hour, which would make the arrays competitive with coal or gas technologies. The company relies on a transfer-printing technique developed by Rogers to peel tiny, high-efficiency solar cells off gallium arsenide wafers and put them onto arrays (J. Yoon *et al. Nature* **465**, 329–333; 2010).

For all the razzle-dazzle of his ideas, however, Rogers in person is laid-back, agreeable and modest. It is telling that he is spending his MacArthur and Lemelson–MIT prize money not on himself, but on supporting students whose ideas might be too risky for government grants. “Life is short,” he says. “I want to be able to point to a few things where we were able to have an impact.” ■

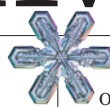


JOHN ROGERS

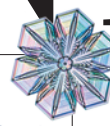
COMMENT

PUBLISHING Rising number of retractions argues for the expansion of peer review **p.449**

TECHNOLOGY Carl Zimmer reflects on the year of the electronic book **p.451**



PHYSICS The delights of making snowflakes in the lab **p.453**



OBITUARY Lynn Margulis, the bold pioneer of early cell evolution **p.458**



ELLIOT D. WOODS/REDUX/EVERETT



The mass protests in Egypt this year were dubbed a 'Twitter revolution' because of the way they were broadcast around the world through social media.

The new history

The past year has shown how prone a highly networked society is to abrupt change. The future of our complex world, says **Philip Ball**, depends on becoming resilient to shocks.

In the 1990s, political scientist Francis Fukuyama, now at Stanford University in California, predicted that the world was approaching the "end of history"¹. Like many smart ideas that prove to be wrong, Fukuyama's thesis was illuminating precisely because of its errors. Events of this year help to reveal why: they were shaped by the unprecedented complexity of today's sociopolitical systems.

After the Soviet Union collapsed, Fukuyama argued that liberal democracy would be the logical and stable end point of civilization. Yet this prospect looks more remote today

than it did at the end of the twentieth century. Protest movements in the fallout from the financial crisis have legitimately challenged the economic system on which recent liberal democracies have been based. The grave debt crisis in Greece has prompted questions over the usefulness of democracy's ultimate tool, the national referendum, and Italy's economic woes have prompted the appointment of a non-elected prime minister. The political situations in Russia and Turkey suggest that democratization is not inexorable, and North Africa and the Middle East show that it can even inflame conflict. Meanwhile, China

continues to show that aggressive capitalism depends on neither liberalism nor democracy. As a 2008 report from the US National Intelligence Council admits, "the Western model of economic liberalism, democracy, and secularism, which many assumed to be inevitable, may lose its luster"².

The real shortcoming behind Fukuyama's thesis, however, is not his faith in the ascendancy of democracy, but his assumption that history is gradualist: that tomorrow will not differ much from today. Instead, the common talk among political analysts now is of discontinuous change³. Discontinuities have ►

► always been with us, in the shape of wars and economic crises, for example. But today, sudden disruptive events don't just happen against a slowly evolving social backdrop — they are woven into the fabric of global affairs.

Take the terrorist attack on the World Trade Center's twin towers in New York in 2001. Many said that this proved Fukuyama wrong — but ten years later, we can see more clearly in what way this was so. It was not simply that the attack was a significant historical event — Fukuyama never claimed that those would cease. Rather, it was a harbinger of a new world order, which the subsequent 'war on terror' failed catastrophically to acknowledge in its attempt to establish conventional battlegrounds in Afghanistan and Iraq. Not only were those wars in no sense 'won', the conflicts were barely wars at all. Armed conflict is no longer corralled from peacetime by national declarations and treaties. It is now a diffuse, nebulous affair: a network of insurgents, militias, terrorist cells, suicide bombers and overlapping and sometimes competing 'enemy' organizations. We have had to say farewell to war and peace.

INTERNAL FEEDBACK

The nature of discontinuous change is often misunderstood. It is sometimes said — this is literally how traditional economists defend their failure to predict the ongoing financial and national-debt crises — that no one can be expected to foresee such radical departures from the quotidian. They emerge, like a hijacked aircraft, out of a clear blue sky. Yet social and political discontinuities are rarely, if ever, random in that sense, even if their immediate triggers have a certain arbitrary character. Rather, they are abrupt in the same way, and for the same reasons, that phase transitions are abrupt in physics. In complex systems, including social ones, discontinuities don't reflect profound changes in the governing forces; they derive from the interactions and feedbacks between the component parts. Discontinuities are therefore precisely what you would expect if you consider today's societies from a complex-systems perspective.

Experience with natural and technological complex systems teaches us that highly connected networks of strong interactions, such as power grids, food webs and computer networks, have a propensity for catastrophic failure and systemic rupture⁴ — in short, for discontinuous change. So it should come as no surprise that today's highly networked, interconnected world is prone to abrupt changes in course.

This phenomenon can be discerned in the revolts in Tunisia and Egypt, dubbed Twitter revolutions because of the way that unrest and news of demonstrations were spread through social networks. Although saying that Twitter caused the revolts is an over-simplification, the networking clearly enabled a random

event to trigger a major one. The Tunisian revolt was set in motion by the self-immolation of a street vendor, Mohamed Bouazizi, in Sidi Bouzid, in protest at harsh treatment by officials. Three months earlier, a similar thing had happened in the city of Monastir — but few people knew about it because it was not publicized on Facebook.

It was surely not without reason that both the Tunisian and the Egyptian authorities shut down Twitter and Facebook. The issue is not so much that these media 'caused' the revolutions, but that they — and their concomitant potential for mobilizing the young and educated — can alter the way things happen in North Africa, the Middle East and beyond. The same tools are now vital to the 'Occupy' protests that are disrupting complacent financial districts worldwide, from New York to

"Social discontinuities are abrupt in the same way, and for the same reasons, that phase transitions are abrupt in physics."

Taipei, drawing attention to issues of social and economic inequality.

Social media seem also to have the potential to facilitate qualitatively new collective behaviours, such as the riots during the summer in the United Kingdom. These brief, destructive paroxysms are still an enigma. Unlike previous riots, they were confined neither to particular demographic subsets nor to areas of serious social deprivation. They had no obvious agenda, not even a release of suppressed communal fury — although they must surely have been linked to post-financial-crash austerity policies. Interaction and feedback, not pre-meditation, seem to lie at their core.

WITHSTANDING SHOCKS

What these events really point to is the profound effect of globalization. They show how deep and dense the interdependence of economies, cultures and institutions has become, in large part thanks to information and communication technologies. And with this transformation come new, spontaneous modes of social and political organization, from terrorist and protest networks to online consumerism, that are especially prone to discontinuous change. Nothing that fails to take this new interconnectedness into account will work: not the economy, not policing, not democracy.

Such extreme interdependence makes it hard to pinpoint, or even to meaningfully define, the causes of major events. The US subprime-mortgage problem caused the financial collapse only in the way that Bouazizi's self-immolation caused the Arab Spring — something else could equally have set the events in motion. The real vulnerabilities were systemic: webs of dependence that had become

destabilized by, say, runaway profits in the US banking industry, or rising food prices in North Africa. This means that potential solutions must lie within such webs too.

Complex systems can rarely, if ever, be controlled by top-down measures. They must be managed instead by guiding the trajectories from the bottom up⁵. In a simple example, traffic lights tend to direct flows more efficiently if they are given adaptive autonomy and allowed to self-organize their switching, rather than following a rigid, supposedly optimal sequence⁶. The Internet can survive random server failures without a disruption to service precisely because no one designed it — it grew its own adaptive topology.

This does not imply that political interventions are doomed to fail, just that they must take forms different from those commonly used today. "Complex systems cannot be steered like a bus," says Dirk Helbing of the Swiss Federal Institute of Technology (ETH) in Zurich, a specialist on the understanding and management of complex social systems. "Attempts to control the systems from the top down may be strong enough to disturb its intrinsic self-organization but not strong enough to re-establish order. The result would be chaos and inefficiency."

But these systems are still manageable, Helbing insists — not by imposing structures but by creating the transactional rules needed to allow the system to find its own stable organization. Healthy financial markets can't be imposed by central management, for example: they must be grown by finding rules that promote systemic robustness.

That's why theories or ideologies tend to be less effective than scenario modelling at predicting or averting crises. It is why problems need to be considered at several hierarchical levels, probably with multiple, overlapping models. And it is why solutions must have scope for adaptation and flexibility.

Cascading crises and discontinuous changes may be unpredictable, but the connections and vulnerabilities that permit them are not. Planning for the future, then, might not be so much a matter of foreseeing what could go wrong as of making our systems and institutions robust enough to withstand a variety of shocks. This is how the new history will work. ■

Philip Ball is a writer based in London.
e-mail: p.ball@btinternet.com

1. Fukuyama, F. *The End of History and the Last Man* (Free Press, 1992).
2. National Intelligence Council *Global Trends 2025: A Transformed World* (US Government Printing Office, 2008).
3. Handy, C. *The Age of Unreason* (Harvard Business School Press, 1990).
4. Vespignani, A. *Nature* **464**, 984–985 (2010).
5. Helbing, D. (ed.) *Managing Complexity: Insights, Concepts, Applications* (Springer, 2008).
6. Lämmer, S. & Helbing, D. *J. Stat. Mech.* P04019 (2008).



The paper is not sacred

Peer review continues long after a paper is published, and that analysis should become part of the scientific record, say **Adam Marcus** and **Ivan Oransky**.

Two months after we started a blog that tracks scientific retractions — Retraction Watch — in 2010, one of us (A.M.) told *The New York Times* that we weren't sure we would have enough material to post with any regularity. That concern turned out to be unfounded — in just 16 months, we have written about some 250 retractions. Little did we know that, in scientific publishing, 2011 would become the Year of the Retraction.

Here's what grabbed everyone's attention: retractions have increased 15-fold over the past decade, while the number of papers has risen by less than 50% (see *Nature* **478**, 26–28; 2011). It is not clear why, and it is always dangerous to draw too many conclusions from what is a relatively rare occurrence — some 300 retractions among 1.4 million papers published annually. About 90 retractions, after all, have come from one author, Joachim Boldt, a German anaesthesiologist, largely because he failed to obtain the appropriate approvals for his research.

Still, it is clear that software that detects plagiarism has played a part in the retraction spike, as has the larger number of eyeballs on papers, thanks to the Internet. It is important

to point out that an increase in retractions isn't necessarily a bad thing, because they correct the scientific record. But the greater visibility of papers and retractions today adds to the evidence revealing why editors need to handle retractions more transparently. In turn, researchers need to stop emphasizing the paper so much.

What is needed, instead, is a system of publication that is more meritocratic in its evaluation of performance and productivity in the sciences. It should expand the record of a scientific study past an individual paper, including additional material such as worthy blog posts about the results, media coverage and the number of times that the paper has been downloaded.

The expanded model would also make it crystal clear to readers when a paper has been corrected or retracted, and why. This would start with better notices from journals to explain those changes. Take the retraction¹ of a paper earlier this year in *The EMBO Journal* by immunologist Silvia Bulfone-Paus. Bulfone-Paus was at the centre of a misconduct scandal at the Research Centre Borstel in Germany, where she stepped down as lab head under pressure in 2010.

The Borstel board found her “ultimately responsible” for the misconduct in her lab and for failing to deal with it in a timely and open manner. (Bulfone-Paus has made few public statements about the case, but she has noted that her results were confirmed by other researchers.) In 2011, journals retracted 13 of her published articles, the stated reasons varying from detailed explanations such as “evidence of data manipulation in Fig. 2C, 4B, and 9, a clear violation of ASM’s ethical standards”², to the wholly unhelpful “This article has been withdrawn by the authors”.

Lines like the latter make us want to pull out whatever hair we have left on our heads. Journal readers should find them similarly frustrating. But we singled out this particular notice for concern not because it said too little, but because, in our view, it allowed the authors to say too much.

TOO LAISSEZ-FAIRE?

The EMBO Journal’s notice¹ also included this: “The authors declare that key experiments presented in the majority of these figures were recently reproduced and that the results confirmed the experimental data and the conclusions drawn from them.”

The statement from Bulfone-Paus and her colleagues described new data and signalled to readers that they could still rely on the original paper, even though it had been retracted. It suggested that the journal stood behind the statement. But when we asked the editor whether that was the case, we were told: “We did not formally investigate this case at the journal and we have not seen this data, as it does not affect the retraction.”³

We've seen a similar lack of close editorial review in correction notices, too. Two recent corrections in *Nature*, and one in *Nature Medicine*, which can only be described as massive, describe in painful detail the number of errors in the original papers. In one, images were improperly labelled and cropped, requiring a solid page of text to explain the changes and how they affect the paper, while another acknowledged that images had been manipulated, which was “not acceptable”.

One of those correction notices, published on 28 September of this year, included this line: “We have also included results from a new, reproduced experiment recently performed with an additional cohort of animals that shows exactly the same results.”⁴ Including new data in a correction notice seemed unusual, so we wanted to know if that line had been subject to peer review. As we reported on our blog⁵, the journal wouldn't say, responding only that peer review is confidential, and that we should talk to the authors — who never responded to our requests for comment.

We don't mean to question the claims in

these particular notices, and we appreciate the arguments for keeping the peer-review process confidential, but we believe it should be no secret whether something has been peer reviewed. Any publishing scientist would surely want journals to assure their readers that vigorous peer review is occurring. After all, peer review is not merely a deterrent, like hydrogen bombs, but an essential element of quality control for journals and for the research community writ large. If it doesn't occur, we think journals owe it to readers to say so, and to explain why. An affidavit from the authors is insufficient.

Requiring any new findings discussed in a correction or retraction notice to be peer reviewed is one step on the route to keeping the scientific record as up-to-date as possible. We believe that some editors also need to work to ensure that their retraction notices say why a paper has been withdrawn, rather than just the exasperating "this paper was retracted by the authors" that we see so often.

Editors have many reasons to pay more attention to retraction and correction notices. For one, scientists often cite papers after they've been retracted, and a clear, unambiguous note explaining why the findings are no longer valid might help to reduce that. But, more importantly, a vaguely worded note that includes further claims from researchers whose work has been seriously questioned, in turn raises questions about the integrity of the journal itself, and about the overall scientific record.

POST-PUBLICATION REVIEW

An even more important step for boosting the long-term credibility of the scientific record is for journals — and scientists — to embrace post-publication peer review. We saw glimmers of this new world after *Science* published the 'arsenic life' paper⁶. Bloggers such as biologist Rosie Redfield attacked the paper⁷, and journalist Carl Zimmer interviewed a dozen experts who had sharp criticisms⁸. But NASA, who employs the lead author, Felisa Wolfe-Simon, refused to engage in the debate until *Science* published a compilation of letters to the editor on the subject (see *Nature* **474**, 19; 2011). (Wolfe-Simon, for her part, has said that the critics are misinterpreting her group's paper.) Responding to critiques in real time has not quite gained widespread acceptance, but many scientists were forced to sit up and take notice of how the scientific record of a paper is expanding.

True, the current system does already allow for critiques. There are letters to the editor, which are robust but also limited by space and often slow to appear. There are online comments on papers, but hardly anyone uses them. Even when scientists do comment, many journals refuse to investigate anonymous criticisms, a policy we've argued against elsewhere. (We applaud the fact that

Nature does look into such critiques.) Faculty of 1000, in which experts flag important papers in their field, is another approach to post-publication peer review.

But these methods are scattered, and there is no reasonable way for scientists to have them all in hand when they're citing a paper.

These developments are why CrossMark (www.crossref.org/crossmark), soon to be launched by CrossRef — a collaborative agency formed by publishers — is so promising. The idea is for every piece of content to include a clickable logo that will let a reader know whether there have been any corrections, retractions or other revisions. It is a solution to the fact that such changes are at best difficult to find — and are sometimes not mentioned at all on 'current' versions of papers.

That is the 'Status' tab on CrossMark. But the platform will also have a 'Record' tab that gives publishers a way to take the idea even further. They will be able to include material they didn't produce, such as blog posts, media coverage, letters, additional data and metrics such as downloads.

This does not mean the end of journals. In fact, it could strengthen the value and extend the imprimatur of those journals that are willing to embrace these new tools, allowing them to isolate the useful notes from the

"It should be no secret whether something has been peer reviewed."

cacophony of what is available, and judge the value of a particular post-publication contribution. Readers will reward those value judgements, passing to their colleagues those

papers with additional content that validates and expands on the results, rendering them particularly trustworthy. If journals aren't willing to start reviewing and compiling additional content related to their papers, someone else will do it.

All of this may mean fewer retractions, because editors would no longer feel limited to such a blunt instrument. We see many papers retracted now that may not need to be, but they contain some nuances that editors don't know how to handle. In the new system, a fleshed-out addendum, or correction, could suffice if the paper included some of the post-publication discussion. We would hope to see fewer cases of the sort that happened this autumn at the Elsevier journal *Genomics*: an "un-retraction" of a study. The journal resurrected a paper that it had withdrawn for authorship reasons, and nothing to do with the substance of the paper⁹. And publishers (we hope) would no longer issue retractions for their own errors, such as running the same study twice¹⁰.

Such a decline would be bad for business at Retraction Watch, but we would be happy if it meant that the scientific record

had become more self-correcting.

We realize that diminishing the importance of the scientific paper will require universities and funding agencies to come up with new ways to judge researchers' productivity. This is a change that most scientists should find heartening. After all, tenure and grant decisions rely heavily on current publication metrics — a flawed system that doesn't reflect how science works. Many people, including *Nature* editor-in-chief Phil Campbell and Cameron Neylon, a senior scientist at the UK Science and Technology Facilities Council, have already begun thinking about how to give credit for contributions other than papers, such as depositing data and writing software or especially worthwhile critiques (see *Nature* **469**, 286–287; 2011).

There are other hurdles. How should scientists treat papers that are hardly read, so are never evaluated post-publication? Does a lack of comment mean that the findings and conclusions are extremely robust, or that no one has cared enough to check? Including readership metrics alongside comments should help here.

None of these issues, however, should stand in the way of taking the crucial steps to make the scientific record more self-correcting. And it is possible now, for the first time, because of the power of the Internet. As a blog that attracts 150,000 page views a month and has tapped into a community of scientists that wants to keep the scientific record up-to-date, Retraction Watch is evidence of how science publishing has changed. It is time to change it further. ■

Adam Marcus and Ivan Oransky
are the creators of Retraction Watch
(retractionwatch.com).

e-mails: ivan-oransky@erols.com;
adam.marcus1@gmail.com

1. *EMBO J.* **30**, 627 (2011).
2. *Mol. Cell. Biol.* **31**, 1330 (2011).
3. Retraction Watch. In retraction notice, Bulfone-Paus "declares" data and conclusions confirmed; journals accept six more retractions. (2 February 2011); available at <http://go.nature.com/ugld7y>
4. Sentürk, A., Pfennig, S., Weiss, A., Burk, K. & Acker-Palmer, A. *Nature* **478**, 274 (2011).
5. Retraction Watch. You can do that? A massive correction in *Nature*, but no retraction. (28 September 2011); available at <http://go.nature.com/arwqj>
6. Wolfe-Simon, F. et al. *Science* **332**, 1163–1166 (2011).
7. Redfield, R. Arsenic-associated bacteria (NASA's claims). (4 December 2010); available at <http://go.nature.com/xcwlys>
8. Zimmer, C. Of arsenic and aliens: What the critics said. *Discover Magazine* (8 December 2010); available at <http://go.nature.com/k3i8mt>
9. Retraction Watch. Retracted retraction leaves *Genomics* paper intact — but authors wonder if anyone will know. (26 October 2010); available at <http://go.nature.com/rxtn75>
10. Retraction Watch. What's going on at Haematologica? Journal double-publishes three different studies. (13 October 2010); available at <http://go.nature.com/rtrazz>



TECHNOLOGY

Rise of the e-book

Carl Zimmer charts the boom in electronic publishing and what that spells for wood pulp and ink.

In the summer of 2010, on a tiny island off the coast of Maine, I saw the future of books. I had been invited to teach a writing course at Shoals Marine Laboratory on Appledore Island, a beautiful bulge of rock covered in scrub and herring-gull nests. During a break at the beach with my family, my wife finished reading her book with typical supersonic speed. She craved another, so decided to experiment with her new iPhone.

She tapped the screen. In seconds, an e-book had streamed invisibly through the air into her hand. Swiping her thumb like a windshield wiper, she soon finished it. She tapped the screen for another. Out of the ether, another e-book appeared.

Now I see, I thought. Everything was in place for a revolution in how we read and write. And the pace of that revolution has surpassed my expectations. Since Apple launched its iBooks application in April 2010, some 180 million books have

been downloaded. Analysts estimate that Amazon will have sold 314 million e-books for the Kindle in 2011 alone. The radical change extends far beyond sales volume: the e-book ecosystem allows writers to reach readers in ways that did not exist before.

Before that moment on Appledore, I was an e-book sceptic. In the 1990s, I got sick of all the promises that the age of e-books was almost upon us — which tended to come from people who wanted you to pay US\$1,000 to come to their ultra-exclusive publishing seminars. When the dot-com bubble popped, the e-book prophets and their pricey seminars disappeared.

It would take another decade for e-books to grow into something more than hype.

➤ **NATURE.COM**
Carl Zimmer learns
to love science films:
go.nature.com/geo44i

People became accustomed to reading on screens. The Kindle, iPad and other tablets made the

experience comfortable and portable. Yet the technology alone did not change how people read. They needed something to read. The first online books, offered by pioneering websites such as Project Gutenberg, were typically the bare-bones text of old books that were already in the public domain. Publishers are now pouring their new books and backlists into e-book stores.

This e-book ecosystem has profoundly changed how we can read books. Volumes that were once out of reach are now wonderfully close. I recently wanted to read *The Hand, Its Mechanism and Vital Endowments as Evincing Design*, written in 1833 by the Scottish anatomist Charles Bell. A few years ago, the only way for me to get my hands on the book would be to go to a university library and hunt it down in the stacks. I did not have to leave my chair to download a free copy from Google Books to my phone, in fully searchable form.

E-books are also changing the experience of writing books. Not long ago, writers had to funnel their books through publishing companies. Now a writer can simply upload a manuscript to websites such as Smashwords, Lulu and Amazon. It can go on sale as an e-book in a matter of hours.

A NEW GENRE

Freed of old constraints, e-books can take on new forms. A writer would never propose a 30-page book to a traditional publisher. Yet many authors are now experimenting with this miniature genre. Some are participating in a programme recently launched by Amazon, called Kindle Singles, to promote these pieces. Writers can also take advantage of the computer power of tablets. One excellent example is *Before the Swarm*, an e-book published earlier this year by The Atavist, a small publishing house. The book is a profile of entomologist Mark Moffett by writer Nicholas Griffin. You can read *Before the Swarm* as a straight profile, or you can tap the screen to reveal what The Atavist calls 'inline extras' — digital footnotes that present audio recordings, video and information such as the definition of the Schmidt Sting Pain Index.

Other science e-books have metamorphosed so far that they are barely books at all. The Elements is an app for the iPad that presents a collection of images and information for every entry in the periodic table. Journey to the Exoplanets, from *Scientific American* and Farrar, Straus and Giroux (both owned by the same parent company as *Nature*), is a combination of text, sumptuous paintings, photographs from space probes, an explorable Solar System and experiments children can run themselves.

E-books not only allow for new experiments, they also make it possible for writers to reach niche audiences. If an acarologist wants to publish a book about the ticks ►

► of New England, he or she does not have to persuade a publisher that it will sell well enough to justify the cost of editing, designing, printing and shipping it. The author can just upload the book themselves.

For some writers, the economics of e-books are attractive. A traditional publisher typically pays the author a royalty of 10–20% on the sale of each book. The sale of international rights can bring more income — but only after an author's royalties rise above the advance. Although e-book distributors such as Amazon or Smashwords won't offer an advance, their royalty arrangements are usually much more favourable. Depending on the e-book publisher, royalties can range from 30–70%.

Of course, it is also true that 70% of zero is zero. A common illusion among writers is that as soon as they finish a book, the world at large suddenly becomes aware of it and millions of copies just sell themselves. In fact, it is easy for a book to vanish into oblivion, and it becomes all the more easy as the number of book titles published each year goes up. Traditional publishers may have their flaws, but they also know how to distribute and publicize books. If writers want to self-publish, they have to take on all the things that publishers do. And today, mastering the art of book publicity is tricky. Many newspapers are shutting down their book-review sections. Discussions about new books are migrating instead to blogs and social media such as Facebook and Twitter.

The format of the e-book itself should also be a cause for concern for authors. Even the most successful e-book campaign simply pushes data from one computer to another. No physical object ends up sitting on a shelf. The longevity of e-books remains uncertain, depending as it does on the technology for reading them. When I look at some of the most elaborate e-books, I hear a ghostly voice whispering, “CD-ROM”. In the early 1990s, compact discs were all the rage — you could fit an entire encyclopedia on a single disc.

For a fleeting moment, CD-ROMs were the future of books. If I had decided to abandon print books and publish my books only on CD-ROMs, I would have imprisoned them in obscurity. Sneer at printed books if you will, but you can't deny that their operating system will never expire. ■

Carl Zimmer is the author of ten books about science, including two e-books: *Brain Cuttings* and *More Brain Cuttings*. e-mail: mail@carlzimmer.com



A food-lab creation: pistachio mousse impregnated with roasted white chocolate.

FOOD SCIENCE

With pipette and ladle

From stretchy ice cream to wire-brushed crackling, **Harold McGee** digests an eclectic modernist menu.

In 1988, the Oxford physicist Nicholas Kurti and his wife Giana published an anthology about food and drink by fellows of the UK Royal Society. When he sent me a copy of *But the Crackling is Superb*, he inscribed it: “An unashamed frivolity!”

The title was a self-deprecating joke, a prominent French chef's polite reaction to an experimental pork roast of Kurti's. Cooking and eating were not respectable academic subjects in those days, so the whole enterprise had the fresh, recreational air of scientists on holiday. The book is still great fun to browse.

Times have changed. Last year, Harvard University in Massachusetts had to turn away hundreds of students from a course on the soft-matter science of cooking. This spring saw the publication of the six-volume treatise on food science and technology in the professional kitchen, *Modernist Cuisine* (The Cooking Lab, 2011). And now we have the serious and substantive anthology *The Kitchen as Laboratory*, edited by food scientists César Vega, Job Ubbink and Erik van der Linden.

Most of the book's 33 chapters are by scientists at universities and food companies, and there is much of interest to cooks and food lovers. The most engaging chapters — the ones that made me feel like getting into the kitchen to experiment — illuminate the making of particular foods and go on to suggest improvements and new twists on them.

In a playful nod to Kurti's anthology (cited by the editors as an inspiration) and to his work in low-temperature physics, *Modernist Cuisine* co-authors



The Kitchen as Laboratory: Reflections on the Science of Food and Cooking
EDITED BY CÉSAR VEGA, JOB UBBINK AND ERIK VAN DER LINDEN
Columbia University Press: 2012.
400 pp./320 pp.
\$29.95/£19.95

Christopher Young and Nathan Myhrvold describe how to achieve the ultimate crackle in duck skin. The keys: a wire pet brush, a water bath and a block of dry ice.

Vega's chapter on soft-cooked egg yolk is a model of clarity and consideration. He demonstrates that yolk viscosity increases with holding time at around 60 °C, gives a chart to help obtain the texture desired, and a table for translating Pascals-seconds of viscosity into the consistency of

mayonnaise, honey or Marmite.

Chemist Martin Lersch surveys the flavour- and colour-forming Maillard reactions, and reports that chopped onions brown faster and get sweeter when fried if you raise their pH with sodium bicarbonate. Alas, he resorts to the most old-fashioned cook's measure, specifying a “pinch” of soda per onion and, confusingly, per half an onion. My evidently oversized pinch browned my onions faster, but gave them a distinctly unpleasant flavour.

Physicist Thomas Vilgis explains how xanthan gum, a readily available bacterial polysaccharide, thickens ketchup and other foods without cooking. He gives a simple recipe that uses it to hold diced vegetables together in a tart. John R. Mitchell, a food

M. HARLAN TURKELL/HARLANTURK.COM

GO NATURE.COM
For a review of
Modernist Cuisine:
go.nature.com/jjxa2t

technologist, follows with an analysis of why starch and gelatin remain better at releasing flavours than xanthan and other thickeners.

There is a substantial report on the Turkish delicacy *salep dondurma* — sometimes known as fox testicle ice cream. It contains the ground roots of an endangered orchid, which are said to resemble fox testicles and give a stretchy, chewy consistency to the ice cream. A group at New York University shows that these properties come from the carbohydrate glucomannan, also found in Japanese konjac flour, which, unlike the orchids, is plentiful. Food-structure expert Tim J. Foster reveals that the stretchiness comes from phase separation of the glucomannan and milk proteins during freezing, a physical effect that can also be achieved by cooking. Expect to see stretchy savouries in modernist restaurants soon.

Other chapters propose making meringue-like solid foams from milk powder and tomatoes, flavouring ice cream with coffee extracted into butter to exclude its bitterness and acidity, and trapping aromas in mixtures of molten sugars. Intriguing but all too sparse are essays from chefs who explain how science contributes to their creativity.

However, I found the book somewhat parochial and not always an enjoyable read. The writing is often less than fluent, and frequent pronouncements about the value of food science become mind-numbing. Some chapters reinforce the image of the food scientist as a technician out of touch with cooks.

In rebutting writer Michael Pollan's denigrations of food science and technology, for instance, one chapter makes the astonishing claim that cheap processed foods are responsible for increased lifespans during the twentieth century. In another example, an encapsulation specialist searches for a quick way to make pizza dough because his mother's recipe takes 2–3 hours and he does not "have the luxury of time". He replaces the yeast with encapsulated baking powders, which most cooks can't buy, and notes that the mixed dough must be rested to hydrate — for 2–3 hours. An essay on the acoustic nature of crispness gives a recipe for a frying batter, then suggests that, to experiment with the recipe, the reader can "place the battered fish on the texture analyzer platen and record the force displacement and acoustic output as the probe pierces the batter".

Food scientists have much to contribute to today's dynamic food scene. I hope that those who write sequels to *The Kitchen as Laboratory* will shake the chips from their shoulders, work more closely with cooks and convey more of the joys of understanding and discovery — and eating. Tip: they should start with another big pinch of Kurti. ■

Harold McGee is a food writer based in San Francisco, California.
e-mail: harold@curiouscook.com



Q&A Kenneth Libbrecht

The snowflake designer

For the past decade, physicist Kenneth Libbrecht has been studying how ice crystals form, taking thousands of photographs of their intricate structures. He describes how he grows snowflakes in his lab at the California Institute of Technology in Pasadena, and never tires of tracking the real thing in the far north.

Why study snowflakes?

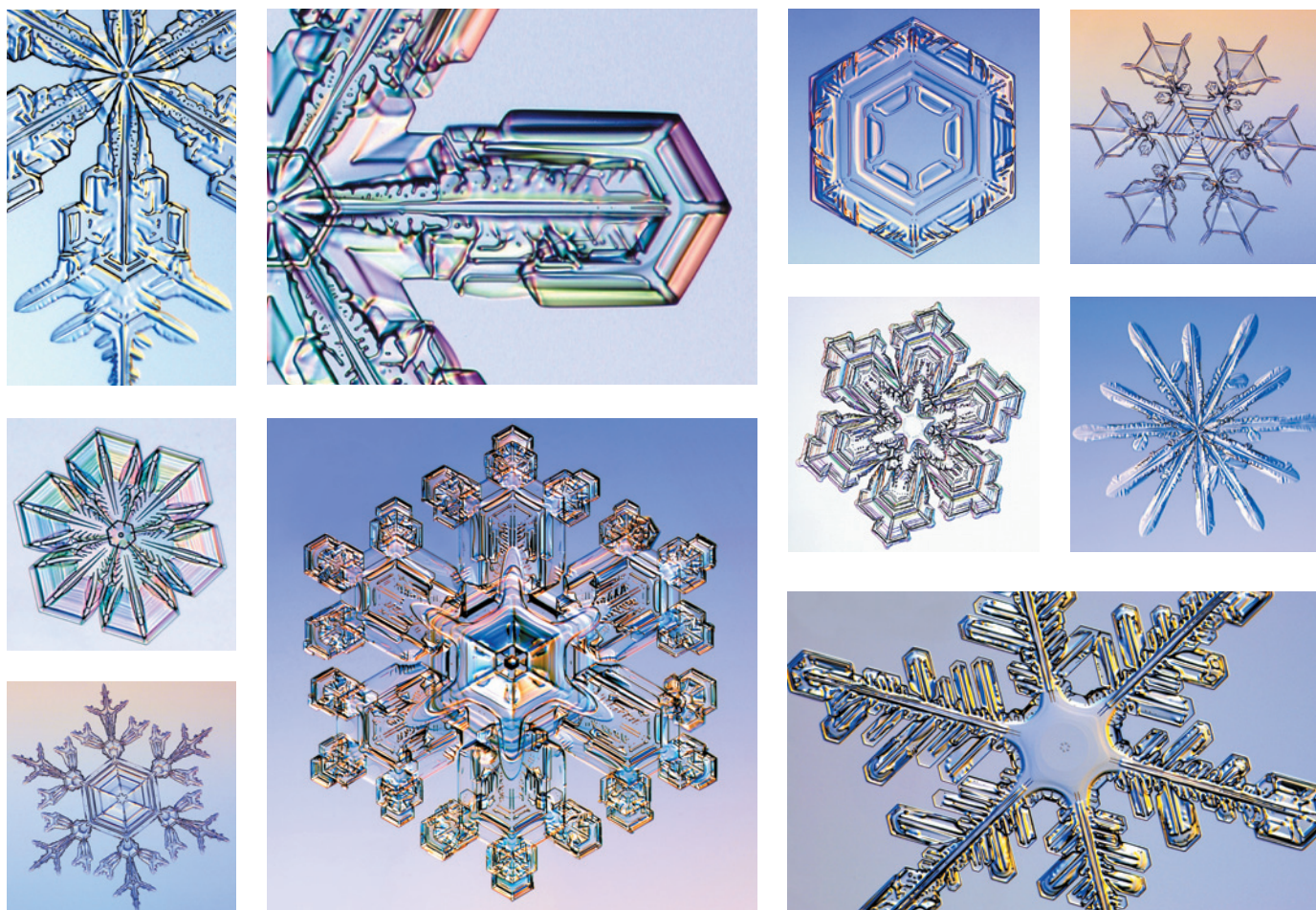
We see these beautiful structures falling from the sky, and we still cannot explain how they came to be. When you ask how snowflakes form, you are really asking about how molecules go from a disordered gaseous state to an ordered crystalline lattice. Unexpected phenomena can emerge — snowflakes are one fascinating example. The complex morphologies



arise in part because different ice surfaces grow at different rates. What we learn could eventually find application in materials science or nanoscale self-assembly. But I am also motivated to simply understand how this natural phenomenon works. I use ice as a case study of crystal growth.

What sorts of shapes do you see?

The diversity of snow-crystal shapes is amazing, and you can learn a lot even with an inexpensive magnifier. You can find hollow columns, needles, bullet rosettes, stellar dendrites, sector plates, ▶



K. LIBBRECHT

Fake or flake? Within this gallery of natural snowflakes are two designed and made by physicist Kenneth Libbrecht in his lab.¹

► 12-branched stars, triangular crystals and many more. One of my favourites is the capped column, which looks a bit like two wheels on an axle. I grew up in snow country in North Dakota, but I never noticed capped columns until I went searching for them. Not every snowfall brings great snowflakes, but some days the crystals are spectacular.

And you're a snowflake photographer?

The physics came first, but I soon became hooked on snowflake photography as well. It is a bit of an odd hobby for someone living in southern California, I admit. My quest for photogenic snowflakes has taken me to Canada, Alaska, Japan, Sweden and locations in the frozen north, yielding more than 10,000 photos. Snowflake photography began in the 1880s with US farmer Wilson Bentley. I am continuing the craft, adding coloured lights and digital imaging techniques.

Why do snowflakes grow into such complex patterns?

A typical stellar snow crystal begins as a tiny hexagonal prism. The six-fold symmetry comes from the underlying symmetry of the ice crystal lattice. Because the six corners of the prism stick out into the humid air, the

corners grow more rapidly, and eventually six branches sprout. The growth is very sensitive to temperature and humidity, which change as the crystal moves through the clouds. The six branches are exposed to the same conditions at the same time, so they grow in synchrony. The final product is a snow crystal that is complex and roughly symmetrical. And because no two snowflakes follow the same path through the clouds, each has a different pattern.

So no two snowflakes can be identical?

I grow simple hexagonal crystals in the lab and many look alike, although they are not identical at the molecular scale. The real question is: how many different ways are there to make a snowflake? With complex patterns, the possible variations can be enormous, easily larger than the total number of atoms in the Universe. So the probability of finding two snowflakes that look identical under a microscope is infinitesimal.

How do you make designer snowflakes?

In a chamber I make millions of little crystals, until one falls flat on a piece of glass. Then I start blowing humid air at it, which makes the crystal grow up in a mushroom-like shape. As it grows, I change the

temperature and humidity to get plates, branches and other desired effects. There is another method using high-voltage electricity. Water molecules are polar, so strong electric fields around a crystal can lead to a runaway instability that produces thin needles of ice. We then grow designer snowflakes on the ends of the needles.

What problems are you working on?

I am especially intrigued by how ice growth changes with temperature. We usually think of snow crystals as thin, flat, star-shaped plates — the iconic snowflake shape. But often the crystals grow into slender hexagonal columns, the same basic shape as a wooden pencil. In fact, the overall shape changes from plate-like at -2°C to columnar at -5°C , then plate-like again at -15°C , then back to columnar below -30°C . This phenomenon was discovered by Ukichiro Nakaya, a Japanese physicist who pioneered the systematic study of snow crystals 75 years ago. But the origin of this odd behaviour is still not known, so I am working hard to solve this puzzle. ■

INTERVIEW BY JASCHA HOFFMAN

¹Artificial snowflakes are at bottom left and top right.

IN RETROSPECT

On the Six-Cornered Snowflake

Philip Ball celebrates the fourth centenary of Johannes Kepler's ice-crystal analysis.

Did anyone ever receive a more exquisite New Year's gift than was given to the German scholar Johannes Matthäus Wackher von Wackenfels, 400 years ago? A booklet of just 24 pages, it was written in 1611 by his friend Johannes Kepler, court mathematician to the Holy Roman Emperor Rudolf II in Prague. In *De nive sexangula* (*On the Six-Cornered Snowflake*), Kepler attempts to explain why snowflakes have their striking hexagonal symmetry.

In his introduction, Kepler writes that he had noticed a snowflake on the lapel of his coat as he crossed the Charles Bridge in Prague, and so came to ponder on its remarkable geometry. This charming, witty work seeded the notion from which all of crystallography blossomed: that the geometric shapes of crystals can be explained in terms of the packing of their constituent particles.

Kepler's analysis of the snowflake comes at an interesting juncture. It unites the older, Neoplatonic idea of a geometrically ordered Universe that reflects God's design with the newly emerging mechanistic philosophy, in which natural phenomena are explained by proximate causes that may be hidden, or 'occult' (like gravity), but are not mystical.

Kepler was not the first to notice that the snowflake is six-sided. That is recorded in Chinese documents dating back to the second century BC, and in the West the snowflake's 'star-like' forms were noted by philosopher and theologian Albertus Magnus in the thirteenth century. René Descartes later explored Kepler's topic, drawing six-fold stars and ice 'flowers' in his meteorological book *Les Météores* (1637), and Robert Hooke's microscopic studies recorded in *Micrographia* (1665) revealed elaborate, hierarchical branching patterns.

"There must be a cause why snow has the shape of a six-cornered starlet," Kepler wrote in *De nive sexangula*. "It cannot be chance. Why always six? The cause is not to be looked for in the material, for vapour is formless and flows, but in an agent." This 'agent', he suspected, might be mechanical: the orderly stacking of frozen 'globules' that represent "the smallest natural unit of a liquid like water" — not explicitly atoms, but as good as.

Here he was indebted to the English mathematician Thomas Harriot, who acted as a navigator for Walter Raleigh's voyages to the New World in 1584–5. Raleigh sought Harriot's advice on the most efficient way to stack

cannonballs on the ship's deck, prompting the ingenious Harriot to theorize about the close-packing of spheres. Around 1606–8, Harriot communicated his thoughts to Kepler, who returned to the issue in *De nive sexangula*.

Kepler asserted that hexagonal packing "will be the tightest possible, so that in no other arrangement could more pellets be stuffed into the same container". This assertion about maximal close-packing is known as Kepler's conjecture. It was proved using computational methods only in 1998, by US mathematician Thomas C. Hales, whose



Mathematician Johannes Kepler explored the snowflake's six-fold geometry in 1611.

proof was published 7 years later (*Ann. Math.* **162**, 1065–1185; 2005). Kepler's vision of crystals as stackings of particles also informed the eighteenth-century mineralogical theory of René-Just Haüy, which forms the basis of all crystallographic understanding today.

Less commonly acknowledged as a source of inspiration for Kepler is the seventeenth-century enthusiasm for cabinets of curiosities (*Wunderkammern*), collections of rare and marvellous objects from nature and art presented as microcosms of the universe. Rudolf II had one of the most extensive cabinets, to which Kepler would have had privileged access. The forerunners of museum collections, the cabinets have rarely been recognized as having any real influence on the nascent experimental science of the age. But Kepler mentions in his booklet having seen in the palace of the Elector of Saxony in Dresden "a panel inlaid with silver ore, from which a dodecahedron, like a small hazelnut in size, projected to half its depth, as if in flower" — a

De nive sexangula

JOHANNES KEPLER

First published 1611.

showy example of the metalsmith's craft that set Kepler thinking about how an emergent order gives crystals their facets.

Yet in the end, Kepler is defeated by the snowflake's ornate form and plate-like shape. He realizes that although the packing of spheres creates regular patterns, they are not necessarily hexagonal or flat, let alone as ramified and ornamented as that of the snowflake. His failure to explain the regularity of the snowflake is no disgrace, for not until the 1980s was this seen to be a consequence of branching growth instabilities biased by the hexagonal crystal symmetry of ice.

Kepler instead fell back on Neoplatonic occult forces. God, he suggests, has imbued the water vapour with a "formative faculty" that guides its form. There is no apparent purpose to the flake's shape, he observes: the "formative reason" must be purely aesthetic or frivolous, nature being "in the habit of playing with the passing moment". That delightful image, which touches on the late Renaissance debate about nature's autonomy, remains resonant today in questions about the adaptive value (or not) of some complex patterns and forms in biological growth.

Kepler signs off his inconclusive tract with an incomparably beautiful variant of 'more research is needed': "As I write it has again begun to snow, and more thickly than a moment ago. I have been busily examining the little flakes."

But the influence of *De nive sexangula* goes further. It was in homage that British crystallographer Alan Mackay called his seminal 1981 paper on quasicrystals *De nive quinquangula*. Here, three years before the experimental work that won Dan Shechtman this year's chemistry Nobel, Mackay showed that a Penrose tiling could, if considered as the basis of an atomic 'quasi-lattice', produce five-fold diffraction patterns.

Quasicrystals showed up in metal alloys, not snow. But Mackay has indicated privately that it might indeed be possible to induce water molecules to pack this way, and quasicrystalline ice was recently reported in computer simulations of water confined between plates (J. C. Johnston *et al.* *J. Chem. Phys.* **133**, 154516; 2010). Whether it can furnish five-cornered snowflakes remains to be seen. ■

Philip Ball is a writer based in London.

Correspondence

LEONARDO DA VINCI/PRIVATE COLL. © 2011 SALVATOR MUNDI/LLC;
PHOTO: T. NIGSWANDER/IMAGINGART



No refraction in Leonardo's orb

Martin Kemp suggests that Leonardo da Vinci's knowledge of optics and minerals is evident in the representation of the orb in the *Salvator Mundi* painting (*Nature* **479**, 174–175; 2011). But I question his interpretation that the double contour of the heel of the hand holding the orb depicts the birefringence (double refraction) arising in a calcite orb.

The painting shows no optical distortion in the folds of the clothes, for example, as would be expected from refraction by an orb of calcite, quartz or glass, or even a water-filled glass vessel. In reality, an inverted and nonlinearly reduced image of most of the chest, arm and shoulder would appear within the orb's outline; the heel of the hand would appear in the top half of this image.

An additional image (nonlinearly enlarged and upright) would open up within the inverted first image if part of the hand or cloth were near the central back surface of the orb. Such refractive effects would be more obvious than any due to birefringence in calcite.

The double contour of the hand continues slightly outside the orb, hence it could be due to a previous stage of the painting, or pentimento. The absence of refraction or reflection effects suggests that the orb depicts an idealized celestial sphere, with

the painted specks on its surface representing heavenly bodies.

André J. Noest *Utrecht University, the Netherlands.*
a.j.noest@uu.nl

Martin Kemp replies:

As far as we can tell, given the damage to the *Salvator Mundi*, the garments behind the sphere are indeed undistorted (*Nature* **479**, 174–175; 2011). But it is wrong to assume that Leonardo da Vinci painted all of the optical phenomena he knew about.

No painters at the time depicted such things as the blur of fast-moving objects or the refraction of limbs in water, because that would have been bad pictorial etiquette. The first blurring in a painting is not seen until more than 100 years later, in the work of Diego Velázquez. Leonardo specifically said that such extreme effects are for the *speculatori* (natural philosophers), not the painters. To show the full effects of the sphere on the drapery behind would have been grotesque in a functioning devotional image.

Contrary to what has become a common belief, Leonardo did not aspire to represent his subjects as if he were a scientist recording natural phenomena. Rather, he was remaking nature synthetically in the functional context of specific paintings (M. Kemp *Leonardo* revised edition, Oxford University Press; 2011).

The birefringence is a side issue and not important for the attribution. The conservator of the picture cannot tell whether the double heel of Christ's hand is deliberate, a result of damage or a pentimento.

Martin Kemp *University of Oxford, UK.*
martin.kemp@trinity.ox.ac.uk

European vultures' altered behaviour

Europe's last remaining populations of griffon vultures (*Gyps fulvus*) in Spain and southern France have taken to

killing livestock, according to the many reports received by authorities. This has provoked discontent and incurred vengeance from some farmers. The alarming departure of the vultures from their normal role as carrion scavengers seems to stem from an increased competition for food resources, which may be caused by changes to European sanitary and conservation policies.

There were 1,165 reported cases of griffon vultures killing domestic livestock in 2006–10 in northern Spain alone, with compensation to farmers costing almost €265,000 (US\$350,000). Unofficial control by poisoning killed 243 griffon vultures in the same period — an ill-conceived action, given that these and other avian scavengers are already severely threatened in Europe.

Changes to European sanitary legislation introduced in 2002 to help combat the spread of bovine spongiform encephalopathy coincided with the introduction of new regulations for animal husbandry, such that any livestock carcasses were collected from farms and destroyed. The combined effects of less carrion and an increase in the griffon-vulture population in Europe may be to blame for the present situation.

Efforts to resolve this conflict are constrained by a lack of solid scientific data, not helped by mounting public alarm and political pressure. Sanitary and other authorities must work with farmers, scientists and conservation groups to produce guidelines to solve the problem in both the short and the long term.

Antoni Margalida *Institute of Ecology and Evolution, University of Bern, Switzerland.*

antoni.margalida@iee.unibe.ch

David Campión *Navarre Environmental Management, Pamplona, Navarra, Spain.*

José A. Donazar *CSIC Doñana Biological Station, Seville, Spain.*

The cell was defined 150 years ago

Anthony Hyman and Kai Simons recount how E. B. Wilson described a cell in 1896 as “the basis of the life of all organisms” (*Nature* **480**, 34; 2011). But it was an almost-forgotten German biologist, Max Schultze, who 150 years ago laid an earlier foundation stone for cell biology by defining the cell in terms of what it contained rather than its boundary.

In an 1861 article, ‘On muscle-particles and what we should call a cell’ (*Archiv für Anatomie, Physiologie und wissenschaftliche Medizin* 1–27; 1861), Schultze rejects the definition of a cell put forward by Robert Hooke almost two centuries earlier.

On the basis of microscopic observations of sections of cork, Hooke in 1665 had introduced the term cell, after the Latin *cella*, for ‘little room with a rigid wall’. Schultze argued that the existence of an enveloping wall, as found in plants, is not an essential criterion for defining a cell.

Schultze based his conclusion on his comparative studies of protoplasmic material from animal muscle tissue and from protozoans. From his observations of these soft, flexible, living systems, Schultze redefined the cell as a “naked speck of protoplasm with a nucleus” (see A. Reynolds *J. Hist. Biol.* **41**, 307–337; 2008).

It could be argued that this more accurate protoplasm–nucleus description of the cell marked the origin of cell biology as a new scientific discipline, encompassing cells as evolving units of all extant and extinct forms of life.

U. Kutschera *Institute of Biology, University of Kassel, Germany.*
kut@uni-kassel.de

CONTRIBUTIONS

Readers are welcome to comment online:
www.nature.com/nature.

Lynn Margulis

(1938–2011)

Biologist who revolutionized our view of early cell evolution.

Lynn Margulis was an independent, gifted and spirited biologist who learned as early as the fourth grade to “tell bullshit from ... real authentic experience”, as she put it in a 2004 interview. With courage, intellect, a twinkle in her eyes and considerable fortitude, she changed our view of cellular evolution.

In 1967, she made the revolutionary case that simple bacteria were incorporated into some early cells to produce the organelles that let plants photosynthesize and animals consume oxygen. This process of endosymbiosis altered Earth by providing plants for animals to eat and additional oxygen for them to breathe.

Her success did not come easily. She had to stand up for her ideas and accept repeated rejections. She submitted her radical paper to a dozen journals before it was published. Remarkably, she developed her theories without training in molecular biology — RNA sequencing was just being developed, and DNA sequencing was years away. Her discovery of the endosymbiotic origins of animals and plants was ahead of the field. She got there on her own terms, and her main insights have been proven correct.

Margulis, who died of a stroke at home on 22 November, was the oldest of four girls born to Leona and Morris Alexander in Chicago, Illinois. She described herself as a bad student who had to stand in the corner a lot. Her parents enrolled her in the University of Chicago Laboratory Schools, with their unique focus on reading original scientific literature. Initially, she didn't like her Lab School and walked out. But after two years back in her local inner-city school, she was glad to return to the University of Chicago. There, she immersed herself in the works of Isaac Newton and Gregor Mendel.

She attributed her scientific success in part to one course: MathSci 2. Biologically oriented, it centred on the material basis of heredity and what connects generations. MathSci 2, she noted, also contributed to the discovery of the structure of DNA — because James Watson also took the course.

When she was 16, while walking up the stairs of Chicago's Ruth Eckerd Hall, she met a tall, handsome, extremely talkative young

man who claimed that he was going to find life in the Universe. That was Carl Sagan. They married when she was 19. She has acknowledged that he had a big intellectual influence on her, claiming gracefully that it would be unfair to say she did it all herself.

The couple went to the University of Wisconsin–Madison, where she obtained

prokaryotes, eukaryotes, mitochondria, and chloroplasts, hot off the press. That paper concluded: “The chloroplasts share a recent ancestry with the blue-green algae”, and that “the mitochondrion shares a recent ancestry with certain respiring and photosynthetic bacteria, the Rhodospirillaceae”. Margulis' proposals for endosymbiotic chloroplast and mitochondrial origins had both been proven in the same paper.

Her research earned her the US President's National Medal of Science, the Darwin–Wallace Medal of the Linnean Society of London, and election to the US National Academy of Sciences and to the Russian Academy of Natural Sciences (as one of only three American members). She could lecture in several languages and was a prolific author and popular speaker. She excelled at balancing research, teaching and raising four children while leading her field. Her second marriage, to crystallographer Thomas Margulis, which had begun in 1967, ended in 1980.

Lynn's sometimes contrary persona was summed up during an interview while recalling her school days at Chicago: “Classes were not required, that's why I went to them all!” She also liked to shock. According to Ann Hirsch of the University of California, Los Angeles, when summarizing bacterial-harboring, nitrogen-fixing root nodules, Lynn referred to them as providing a place for bacterial “food and sex”.

My favourite memories of Lynn were her lengthy calls to kick ideas around (she abhorred e-mail). She was always upbeat and ready to explore. Her visit to our home in 2007 was full of adventure and laughter. She told us about her travels, loved our visit to the Getty Center in Los Angeles, California, and its travertine fossils, and always had a ziplock bag of yerba maté on hand. She'll be greatly missed. ■

James A. Lake is in the departments of molecular, cell and developmental biology, and human genetics, University of California, Los Angeles, Los Angeles, California 90095, USA. He became friends with Lynn Margulis after meeting her at a conference at Woods Hole in Massachusetts in the 1980s. e-mail: lake@mbi.ucla.edu



B. O'CONNOR

her MA in zoology and genetics, and then to the University of California, Berkeley, where she began a PhD in genetics. Their marriage ended before her doctorate was awarded, and Lynn moved to Massachusetts in 1963 with her two sons to take up a biology lectureship at Brandeis University in Waltham. She received her PhD from Berkeley in 1965, and then moved to Boston University, where she remained for 22 years. There, she published her paradigm-changing book, *Origin of Eukaryotic Cells* (1970). In 1988, she moved to the University of Massachusetts, Amherst.

As Boston University's Douglas Zook, then an undergraduate in one of her classes, recalled, it was an emotional moment in 1978 when her ideas on endosymbiosis were confirmed. She strode into class beaming, holding Robert Schwartz and Margaret Dayhoff's classic paper, 'Origins of

TEA WITH JILLIAN

Robotic reflections.

BY BRENDA COOPER

On 25 June 2054, Technical Nurse Paul Castle brought a program he had been working on for three years into Shady Acres Nursing Home. He'd pieced it together from bits of open source available on the web and from some failed research of his own, which he had hoped to turn into a thesis project. He had tested it with crowd-sourced volunteers in Thailand. He'd done it for a patient, and because memory fascinated him.

This morning started like every other. Paul arrived early and perched at his desk, which had a view of the common kitchen for his wing, the long hallway between rooms, and of images from every room in the building. He did this just to watch the most beautiful of the robots in all of Shady Acres prepare Jillian's breakfast. She worked with precision — like all robots — never spilling a drop of the oatmeal, adding exactly the same number of raisins and the same amount of sugar. The robot stirred in a half a cup of milk the same way every morning, and added the appropriate sprinkle of tasteless vitamin powder. Then she poured a glass of faux orange juice and glided down the hallway from the common kitchen to Jillian's suite.

That was the moment Paul thought of as his meditation, his reminder to be as precise as Jillian's robot nurse, as beautiful as he could manage in every interaction with the staff and residents.

There were other robots, of course. Some looked like people. Other residents chose the cheaper and more mechanical option of wheeled bots with screens or air-displays on them and metallic arms and hands for dispensing medications, making food and helping with bedding. These often ended up decorated; his favourite had stuffed golden retrievers tied to the large central post so their heads and ears flopped around as the robot negotiated stairs or tight turns. That one belonged to Patrice Mallo, who had been a good enough dog breeder that she could afford a single-room suite. For her part, Jillian dressed her caretaker in scarves and hats and gloves, and sometimes in evening gowns. On the morning of the 25th, Jillian had dressed her robot in pink.

Jillian owned the Penthouse. She had inherited a great pile of money from a grandfather, but she'd lost her ability to do more than shuffle the halls, and now she needed

help cooking and cleaning and — on some days — remembering her name.

Jillian was the loneliest person Paul had ever met. He stood in for family on visiting days, and spent 20 minutes with her and the robot and Jillian's robotic dog every afternoon at the end of his shift. He had a real dog, and parents to go home to, but just like his day started with Jillian's breakfast, it ended with her cup of tea.

The robot girl would bring in the tea, leaning down and setting the lacquer tray precisely between them. They talked over this tea, small talk about the weather, about Paul's dog Maximus whom he picked up at the end of every day and walked through Central Park. Sometimes they talked about Jillian's past, and when this made Jillian cry, Paul would dry her eyes and ask her why. The most common answer was: "I miss being home. I miss being young and spry and beautiful."

On 25 June, Paul spilled his tea on the table, so that some of the hot liquid splashed Jillian across the shoulder. This gave him an excuse to slip the data pearl from her necklace as he dried it off and add his program to her interface jewellery.

It took two days before he began to see results. The first thing he noticed was a change in the way the robot walked. Her hips slid right and left as she walked. It wasn't quite feminine, but neither was it robotic. He imagined Jillian walking that way when she looked like a fully fleshed version of her metal companion. The idea made him smile.

At tea that day, Jillian looked happier. Her hands still shook as she held her china cup, her orange lipstick still missed the corners of her mouth, and her thin hair still clung to her cheeks. But her eyes were brighter and she gave him a smile that he imagined was just a touch more aware.

Weeks passed.

The robot began to join them for tea, to talk to Jillian about her past in a soft, silky and metallic voice. The two spent more time together. They bent their heads over books. The robot girl watched

vids with the old woman, so close that metal touched skin often enough that Paul had to powder the old woman's legs so she wouldn't be burned by the friction of the robot's movements. Jillian even named the robot after herself, calling her Jilly.

Over tea, Paul spoke softly. "Does it help you when Jilly can keep your memories for you?"

"Yes." She paused. "I like it that when I talk to her she can recall the way the garden smelled after one of Poppa's parties."



"Are you happier?"

"Yes thank you. I know you helped to do that."

He hadn't expected that. "How?"

"Jilly told me. She remembers the day you spilled the tea, and how it felt to have the interface gone and returned, and how more kinds of things I want to tell her get stuck in her head so she can take them out for me later. She says you have made her into my mirror." Jillian took a sip, age-spotted hands shaking so the liquid almost spilled from the cup. "Thank you." ■

Brenda Cooper's latest novel is *Mayan* December, available from Prime Books. Brenda lives in the United States in a family that includes as many dogs as people. See www.brenda-cooper.com for more information.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
www.nature.com/mtoodm

The ultimate fate of planets

A planetary system has been found in a startlingly tight orbit around an evolved star. The finding challenges the idea that close-in planets are destroyed as their host star evolves. [SEE LETTER P.496](#)

ELIZA M. R. KEMPTON

Billions of years from now, the Sun will expand to form a red-giant star. At that time, the Sun's atmosphere will extend beyond the orbit of Earth, and our star will engulf the current orbits of the inner planets of the Solar System — Mercury, Venus and Earth. These three planets are expected to ultimately die a fiery death by being swallowed up by the Sun¹. Yet on page 496 of this issue, Charpinet *et al.*² present an analysis of data from NASA's Kepler satellite that reveals a system of planets that has probably survived such a cataclysmic event.

The authors found two planets in orbit around an evolved star — classified as a hot B subdwarf — that has reached the next rung of the stellar evolutionary ladder beyond the red-giant phase. The two planets' orbits take them extremely close to their host star — at distances of less than one per cent of the Earth–Sun distance. At these distances, the planets would have been engulfed deep within their host star during its time as a red giant. Yet these two planets seem to be alive and well, which begs the question, how did they survive?

Charpinet *et al.* posit that the two planets originally resided farther away from their host star, and that their orbits were subsequently dragged inward during the star's expansion to become a red giant, which ultimately engulfed the planets³. In this scenario, both planets may have originated as massive gas giants, similar to Jupiter. Their gaseous atmospheres would have been stripped away as the planets were immersed by their host star, leaving behind two bare cores of rock when the star shrank back down to dwarf-star stature by shedding its own outer layers (Fig. 1).

The method that Charpinet and colleagues used to detect this planetary system is interesting in its own right. It differs from the usual way in which the Kepler satellite is used to discover planets, which is to measure the light that is blocked out by a planet passing in front of its host star. Here, the host star, named KIC 05807616, is known to be a variable star that pulsates (brightens and dims) at regular intervals^{4–6}. However, when the authors² analysed the light from this star, they found variability at two specific periods that did

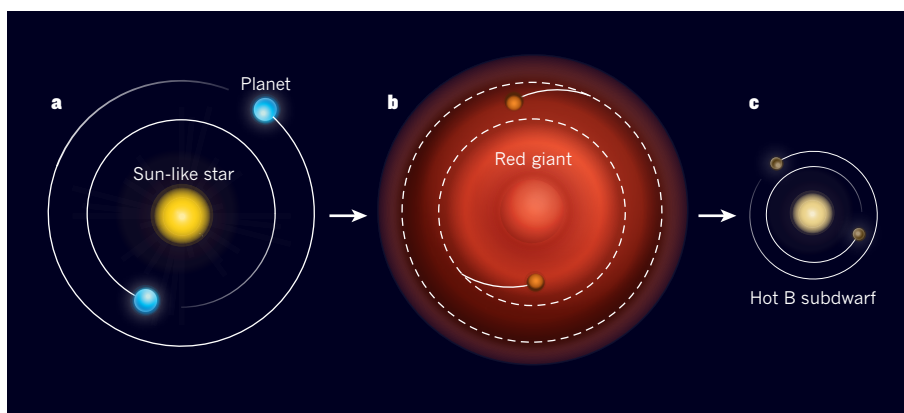


Figure 1 | Possible formation mechanism of close-in planets. **a**, The two planets identified by Charpinet *et al.*² were probably massive gas-giant planets during the early part of their host star's life, when the star burned hydrogen into helium as our Sun does today. At this time, the planets resided farther away from their host star than they do today. **b**, When the star exhausted the supply of hydrogen in its core, it expanded to become a red-giant star. The outer layers of the star then engulfed the orbits of the two planets. The planets lost their outer gaseous layers as their orbits spiralled inwards. **c**, For unknown reasons, the red giant expelled its cool outer layers, leaving behind a small hot B subdwarf star. The small rocky cores of the initial planets were left behind in close-in orbits with periods of 5.8 and 8.2 hours. (Figure not drawn to scale.)

not seem to result from the star's pulsations. Six months of uninterrupted data from Kepler revealed that the light from the star was periodically brightening and dimming every 5.8 and 8.2 hours. But it was unlikely that this variability resulted from the star itself, which pulsates much more quickly. Instead, Charpinet *et al.* floated the hypothesis that the apparent pulsations were due to the presence of two planets in orbit around the star. In this case, the anomalous pulsation periods found in the Kepler data correspond to the orbital periods of the two planets.

The periodic brightening and dimming observed every 5.8 and 8.2 hours results from the changing amount of stellar light that is reflected off the planets as they orbit their host star — changes analogous to the phases of our Moon. Furthermore, light given off directly by these hot planets similarly oscillates with the orbital periods of the planets, further contributing to the periodic signals discovered. The variability that results from these subtle shifts in the planets' reflected and emitted light is tiny. Charpinet *et al.* found that this variability is present at a level of only several hundredths of one per cent, which means that the planets

themselves are probably quite small. Using some reasonable assumptions about how these planets emit and reflect light, the authors concluded that both planets are probably comparable in size to Earth. Light that is directly emitted or reflected from extrasolar planets has been detected in the past, but this is the first time that this particular method has been used for the discovery of a planetary system.

The prospect of planets being detected in close proximity to an evolved star is certainly of great interest, because it was previously assumed that such objects would be destroyed during the star's evolution. But it is important to note that, using their current observations, Charpinet *et al.*² are unable to fully determine either the size or the mass of the two planet candidates. Although these objects are most likely to be approximately Earth-sized, the authors cannot say with certainty that they are not considerably larger (or even smaller). This uncertainty limits their ability to draw firm conclusions about the origin of the planetary system. If the planets are not remnants of the original system, one intriguing alternative possibility is that the original planets orbiting KIC 05807616 were completely destroyed

during the star's red-giant phase. According to this hypothesis, the two close-in planets formed anew, after the star receded from its red-giant phase, from material that was left behind when the star expelled its outer layers.

Surveys that have searched for close-in planets orbiting evolved stars have demonstrated a lack of such objects^{7–9}. And yet Charpinet and colleagues' results buck this trend by revealing a planetary system in a startlingly tight orbit. One way to resolve this apparent discrepancy is to postulate that close-in massive planets (that would have been easily detected by the previous surveys) are almost, but not completely, destroyed during the red-giant phase of stellar evolution: their small, dense cores may survive.

Stars such as the Sun spend billions of years converting hydrogen into helium before undergoing a quick and violent growth phase to become red giants. The expansion of the Sun to its red-giant phase will surely kill off all life on Earth in the process. However, the existence of planets orbiting an evolved star points to an interesting possibility that all close-in planets are not entirely destroyed during stellar evolution. Charpinet and colleagues' results therefore have important implications as we strive to understand the ultimate fate of planetary systems. ■

Eliza M. R. Kempton is in the Department of Astronomy and Astrophysics, University of California, Santa Cruz,

Santa Cruz, California 95064, USA.

e-mail: ekempton@ucolick.org

1. Schröder, K. P. & Smith, R. C. *Mon. Not. R. Astron. Soc.* **386**, 155–163 (2008).
2. Charpinet, S. *et al. Nature* **480**, 496–499 (2011).
3. Villaver, E. & Livio, M. *Astrophys. J.* **705**, L81–L85 (2009).
4. Green, E. M. *et al. Astrophys. J.* **583**, L31–L34 (2003).
5. Charpinet, S. *et al. AIP Conf. Proc.* **1170**, 585–596 (2009).
6. Van Grootel, V. *et al. Astrophys. J.* **718**, L97–L101 (2010).
7. Johnson, J. A. *et al. Astrophys. J.* **675**, 784–789 (2008).
8. Sato, B. *et al. Publ. Astron. Soc. Jpn* **60**, 539–550 (2008).
9. Wright, J. T. *et al. Astrophys. J.* **693**, 1084–1099 (2009).

ECOLOGY

Bleak future for amphibians

The major threats to amphibian species include pandemic disease and changes in climate and in land use. A study of the global distributions of these threats predicts that they will affect most amphibians by 2080. [SEE LETTER P.516](#)

ROSS A. ALFORD

Modern amphibians are the oldest extant group of terrestrial vertebrates, having existed for hundreds of millions of years¹. Some have undergone very little evolutionary change over this period. But amphibians are far from being primitive animals of low diversity. They are, in fact, more diverse than mammals or reptiles — 6,894 species are presently known², and the best estimate³ for the total number of species is 8,000–12,000. They occupy every terrestrial habitat except Antarctica and the high Arctic, and have an important role in nutrient dynamics, in the cycling of energy flows between terrestrial and freshwater systems, and in controlling populations of pest insects¹.

Amphibians are also the terrestrial vertebrates most at risk of extinction: at least 37% are classed as vulnerable, threatened or endangered⁴ (Fig. 1). This figure is certainly an underestimate of the risk, because the conservation status of about 25% of species remains unknown⁴. Sadly, things are likely to be much worse by 2080,

as Hof *et al.*⁵ report on page 516 of this issue.

Ever since amphibian declines were recognized as a global problem⁶, it has been suggested that the phenomenon has several causes. Early explanations focused on pollution, habitat loss and climate change⁷, but many declines have occurred in relatively pristine areas not strongly affected by habitat modification, and far from sources of environmental contaminants. In these cases, analyses

have indicated that climate change alone is unlikely to be the reason⁷.

A major cause of many of these enigmatic declines was discovered⁸ in 1998. The pathogenic fungus *Batrachochytrium dendrobatidis* causes chytridiomycosis, a disease that infects amphibians across a wide range of taxa^{8,9}. This disease has emerged in many areas, and has caused entire regional faunas to decline dramatically, to the point of local extinction. According to the International Union for Conservation of Nature, many amphibian species became critically endangered between 1980 and 2007, in most cases probably because of the global chytridiomycosis pandemic⁷. Habitat loss and climate change have received less attention as causes of amphibian decline, but threaten many more species in the long run⁷. Climate change may already be affecting the course of the chytridiomycosis pandemic, and its influence is likely to continue during this century, because vulnerability to the disease is strongly affected by weather and climate^{7,10}.

Given the current threats to amphibians, it is crucial to develop plans to protect their diversity. This requires a good understanding of present and future dangers. Hof *et al.*⁵ have undertaken the complex task of examining the spatial extent of three of the major threats to the global diversity of amphibians — disease, climate and land-use change. By comparing their findings with the spatial distribution of amphibian diversity, they were able to say how the threats have changed and will probably change in the period from 1980 to 2080. The authors did this for all three orders of modern amphibians, which differ in their levels of ecological and taxonomic diversity: roughly 90% of modern amphibians are frogs, 8% are salamanders and 2% are the



Figure 1 | The edge of destruction. The armoured mistfrog *Litoria lorica* was thought to have been extinct for 17 years, but was rediscovered¹⁴ in 2008. Many other amphibian species are in decline, a situation that Hof *et al.*⁵ predict will accelerate during the twenty-first century.



50 Years Ago

The heavy nuclear explosion on October 30, 1961, at 8.33.33 G.M.T. at a distance of 1,160 km. in Novaya Zemlya (presumably at tropospheric heights) was recorded at Sodankylä by means of a seismograph, a microbarograph, a magnetograph, and a vertical incidence ionosonde. The deflection of the microbarograph took place at 9.42 G.M.T. with an amplitude of about ± 1 mb ... On October 31, the microbarograph again showed two very distinct and strong deflexions, namely, at 18.32 and 21.38 G.M.T. These deflexions are interpreted as being caused by round-the-world waves due to the same nuclear explosion, one being propagated in the backward, the other in the forward, direction. The mean velocity deduced from these round-the-world waves is 311 m./sec ... The waves are supposed to have been guided in the spherical shell between the ground and the stratopause.
From *Nature* 23 December 1961

100 Years Ago

The Rubber-Planter's Notebook. By Frank Braham — This book is what it purports to be, a handy book of reference on Para rubber planting, with hints on the maintenance of health in the tropics and other general information of utility to the rubber planter ... The author's section on general information will be found specially useful ... for the young planter going out to the East for the first time; but for the older resident in the tropics "drink as little as possible—fluids inflate the bowel" is dangerous advice ... If blackwater fever is encountered death in such cases may be the result ... In these essential rules also mention of the all-important hot bath and change at sundown would have added to their completeness.
From *Nature* 21 December 1911

less well-known caecilians¹. Frogs are far more widely distributed than the other groups.

Hof *et al.* used a complex, wide range of modelling approaches in their study. Briefly, they used data on the distributions of 5,527 amphibian species in bioclimatic models to predict the global distribution of the species on a latitude–longitude grid consisting of cells $2^\circ \times 2^\circ$ in size. This analysis took into consideration a broad range of future climate scenarios proposed by the fourth Intergovernmental Panel on Climate Change. To forecast the spread of chytridiomycosis, they used a previously published model¹¹ that predicted the distribution of the causative fungus *B. dendrobatidis*. Their data on land use and land-use changes came from the Millennium Ecosystem Assessment, a report on the current state and the future of Earth's ecosystems.

According to Hof *et al.*⁵, the outlook for amphibians is not good. For frogs — the most diverse group — the areas most affected by climate change coincide with regions of greatest species richness. The authors' models indicate that, in some of the regions with the greatest diversity of frogs, more than half of the species will probably be negatively affected by climate change by 2080. Strong climate-change impacts are also likely for some salamanders, particularly tropical faunas.

The models also suggest that land-use changes, especially in tropical regions, are likely to have strong negative effects on amphibians in some of the areas that have high levels of amphibian diversity. Finally, they predict that the distribution of *B. dendrobatidis*, and thus possibly of chytridiomycosis, will be focused in temperate and mountainous areas. This is better news for frogs, which reach their peak diversity in the lowland tropics, but may be bad news for salamanders, whose centre of diversity is in northern temperate regions.

Possibly the worst news is that, on the whole, the areas most affected by each category of threat do not coincide geographically: less than half of the grid cells in the 25% of land most threatened by any one factor are also in the 25% most threatened by any other factor. Because the threats are spread out, more than half of the total geographic distribution of each major amphibian taxon is in areas that Hof *et al.* predict will be highly affected by at least one of the three threat factors by 2080. The picture becomes worse when only the most diverse faunas are considered — roughly two-thirds of the areas that have the highest diversities of frogs and salamanders are likely to be highly threatened in some way.

The effects of major changes in land use will probably be as strong as, or even stronger than, Hof and colleagues assume, because the complex life histories of amphibians may render them particularly vulnerable to the disruptive effects of habitat modification¹². But in other respects, the exceedingly gloomy picture presented by the authors might turn out

to be too pessimistic. For example, the exact effects of climate change and chytridiomycosis on amphibians are not known, and so their overall impact may be less than is predicted⁵. The somewhat coarse grid used in Hof and colleagues' bioclimatic modelling might also obscure small-scale variations that could allow species to avoid the negative effects of climate change by shifting their habitat ranges relatively short distances, or simply by changing how they use their existing ranges (for example, by choosing less exposed retreat sites)^{5,7}. Moreover, the authors' analysis equates the presence of *B. dendrobatidis* to negative conservation effects of chytridiomycosis, but the impact of the disease varies strongly among regional faunas, ranging from disastrous population collapses in some areas to little or no effect in others^{9,13}.

On the other hand, some of Hof and co-workers' results may be overly optimistic. For example, they did not model possible non-additive impacts of threats, such as the strong possibility that the threat of epidemic outbreaks of chytridiomycosis may worsen with changing climate^{7,10}, or that habitat modification may restrict amphibians' ability to resist climate change by altering their habitat preferences^{5,7}. Nevertheless, their work is a valuable step towards a true understanding of overall threat levels to an iconic group of animals. It is also a sobering reminder of how much critical information is needed before we can truly understand the extent of anthropogenic threats to global biodiversity, or be fully prepared to rationally manage them. ■

Ross A. Alford is at the School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia.
 e-mail: ross.alford@jcu.edu.au

1. Alford, R. A., Richards, S. J. & McDonald, K. R. in *Encyclopedia of Biodiversity* (ed. Levin, S. A.) 1–12 (Elsevier, 2007); <http://dx.doi.org/10.1016/B0-12-226865-2/00013-4>.
2. www.AmphibiaWeb.org (accessed 27 November 2011).
3. Parra, G. *et al.* in *Amphibian Conservation Action Plan: Proc. IUCN/SSC Amphibian Conserv. Summit* (eds Gascon, C. *et al.*) Ch. 10 (IUCN, 2005).
4. IUCN Red List of Threatened Species. Version 2011.2. www.iucnredlist.org (downloaded 27 November 2011).
5. Hof, C., Araújo, M. B., Jetz, W. & Rahbek, K. *Nature* **480**, 516–519 (2011).
6. Blaustein, A. R. & Wake, D. B. *Trends Ecol. Evol.* **5**, 203–204 (1990).
7. Alford, R. A. in *Ecotoxicology of Amphibians and Reptiles* 2nd edn (eds Sparling, D. W., Lindner, G., Bishop, C. A. & Krest, S. K.) 13–46 (CRC Press, 2010).
8. Berger, L. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
9. Lips, K. R. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 3165–3170 (2006).
10. Pounds, J. A. *et al.* *Nature* **439**, 161–167 (2006).
11. Rödder, D., Kielgast, J. & Lötters, S. *Dis. Aquat. Organ.* **92**, 201–207 (2010).
12. Becker, C. G., Fonseca, C. R., Haddad, C. F. B., Batista, R. F. & Prado, P. I. *Science* **318**, 1775–1777 (2007).
13. Daszak, P. *et al.* *Ecology* **86**, 3232–3237 (2005).
14. Puschendorf, R. *et al.* *Conserv. Biol.* **25**, 956–964 (2011).

ATOMIC PHYSICS

When ultracold is not cold enough

A technique for cooling ultracold atoms in optical lattices has been demonstrated. This advance should allow the physics of strongly correlated systems, including that of quantum magnetism, to be explored. SEE LETTER P.500

GRETCHEN K. CAMPBELL

Experiments with ultracold neutral atoms routinely reach nanokelvin temperatures. When combined with optical lattices — light ‘crystals’ created with standing waves of light — ultracold atoms are an almost ideal quantum many-body system^{1,2}. Thus, they are an excellent platform for simulating the physics of solid materials. Lattice-trapped atoms can simulate theoretical model systems that are relevant to understanding strongly correlated materials — systems in which electrons interact strongly. However, for some model systems, even nanokelvin temperatures can be too hot for simulating the relevant phenomena³.

On page 500 of this issue, Bakr *et al.*⁴ demonstrate a technique for cooling quantum

atomic gases in optical lattices that may allow much lower temperatures to be reached than those currently attainable. This opens up the possibility of studying a number of outstanding problems in many-body physics, such as quantum magnetism and high-temperature superconductivity^{1–3}. In addition, the control achievable with this technique may provide a way of producing logic gates based on two quantum bits (qubits) and creating quantum registers — the working memory needed for quantum computing — using ultracold atoms.

Bakr and colleagues’ technique for cooling atomic gases relies on atom blockade. Blockade phenomena arise from the interactions between tightly confined particles. If the interaction energy is sufficiently high, it is much harder to add another particle to the system,

because of the increased amount of energy required. Blockade phenomena are used in many systems. For example, Coulomb blockade of electron charges can be used to make transistors based on single electrons⁵. Blockade effects have also been proposed as a way to produce qubits^{6–8}.

In ultracold-atom experiments, atom blockade occurs as a result of repulsive interactions between the atoms. When trapped in an optical lattice, ultracold atoms develop an energy-band structure just like that of electrons in a solid. The higher the number of atoms in a given lattice site, the higher the interaction energy, creating a barrier to the addition of a further atom. If the optical lattice’s sites are sufficiently deep, these interactions give rise to insulating behaviour, analogous to the insulating behaviour of electrons in some solids. As a result, there are a fixed number of atoms per site and tunnelling of atoms between different lattice sites is inhibited.

In their study, Bakr *et al.*⁴ show that, in addition to this transport blockade, a blockade can occur if atoms are excited to different energy bands within a single lattice site. The authors introduce the phenomenon of orbital exchange blockade (OEB), which allows the entropy (and thus the temperature) of the system to be reduced.

Bakr *et al.* demonstrated OEB by first preparing a two-dimensional gas of rubidium-87

PSYCHOLOGY

Who needs a leader?

In a dance class, everyone follows the instructor. The opposite situation would be if everyone in the class performed without a designated leader — an activity known as joint improvisation. In a paper published in *Proceedings of the National Academy of Sciences*, Noy and colleagues investigate which of two such situations is the more effective (L. Noy *et al.* *Proc. Natl Acad. Sci. USA* <http://doi.org/hbz>; 2011).

Day-to-day examples of joint improvisation include two toddlers playing together. A rather structured example is improvisation in artistic performances (pictured). By the improvisers’ own admission, there are ‘moments of togetherness’ when the level of performance is high but no one knows who is leading. But how does this work? The lack of an experimental paradigm means that this question has not been studied extensively — at least not for open-ended actions.

Noy and colleagues designed an experiment based on the mirror game, a widely used theatrical practice. Specifically, they asked two players each to move a

handle along one of two parallel tracks in one dimension. The instruction was: ‘Imitate each other, create synchronized and interesting motion, and enjoy playing together.’ The nine one-minute rounds in each game were of two types — leader–follower rounds and joint-improvisation rounds. The authors measured the players’ movements with high resolution in time (20 milliseconds) and space (1 millimetre).

They investigated the movements of expert players — artists with more than 10 years of experience in joint improvisation. In the leader–follower rounds, the follower showed jittery motion, which oscillated around the leader’s confident movement. By contrast, with no designated leader the players performed better, reaching lower errors in velocity of movement and stopping times. In fact, the players jointly showed confident motion 12% of time, compared with 2% in the leader–follower situation.



J. RENSTEN/GETTY

So, is having a leader really counterproductive? It depends. With novice players, Noy *et al.* obtained opposite results: these players performed much better with a designated leader. As moments of togetherness are rare in day-to-day life, having a leader is perhaps beneficial for most of us, at least while we learn a new skill. **Sadaf Shadan**

atoms in a square optical lattice created using two perpendicular standing waves of light. Initially, they arranged the system such that there was a known number of atoms per individual site, and they prepared the atoms in the same quantum state — the ground state, or band, of the lattice. If the neighbouring sites were already occupied, the repulsive interactions between the atoms suppressed tunnelling between lattice sites.

Next, by changing the intensity of one of the standing waves, the authors modulated the depth of the lattice sites. If the frequency of this modulation was tuned to exactly match the separation between the ground band and an excited band of the lattice, atoms could be excited to the higher band. In a similar situation to that for suppression of tunnelling between sites, if an atom was already excited to the higher band, the excitation of a second atom in the same site was suppressed as a result of the interactions. Because of this atom-number-sensitive OEB, the frequency needed to excite atoms depends on the number of atoms in the ground and excited bands. Atoms excited to the higher band could then be removed from the system. Bakr *et al.* demonstrated that OEB allows spectroscopic differentiation of sites with different numbers of atoms and control over the final number of atoms in a given site.

Having a lattice-trapped atomic system at finite temperature creates ‘defects’ in it. Instead of a uniform filling, defects exist in lattice sites at which either no atoms are present (holes) or there are extra atoms. OEB cannot be used to directly fill the holes, but by using atom-number-sensitive OEB, sites with extra atoms can be corrected, removing entropy from the system and so cooling it.

Bakr *et al.* achieved such atom removal in two experiments. In the first one, they initially loaded the lattice with atoms such that the system was in an insulating state — the ground state for the system — with a known number of between one and four atoms per site. Next, by sweeping the modulation frequency, they deterministically reduced the lattice occupation number to a single atom per site.

In the second experiment, the authors⁴ loaded the lattice such that the system was not in the ground state. Instead, they loaded a random number of atoms per site, and by sweeping the modulation frequency, they were able to remove all of the ‘extra’ atoms. After the frequency sweep, they adjusted the depth of the lattice sites so that the final state, an insulating state with a single atom per site, was close to the ground state. This cooling ‘algorithm’ is analogous to heat-bath algorithmic cooling, in which entropy is pumped out of a system into a heat bath. Heat-bath algorithmic cooling has previously been demonstrated using nuclear magnetic resonance of solid-state qubits⁹.

Bakr and colleagues’ technique allows entropy to be removed from the system. But it also allows individual lattice sites to be targeted, as has been shown previously¹⁰. The combination of these two features could facilitate the creation of a quantum register consisting of thousands of atoms, and provide a roadmap to scalable quantum computing using lattice-trapped atoms. Achieving picokelvin temperatures (1 picokelvin is 10^{-12} kelvin) for lattice-trapped atoms remains an important goal, which the authors’ technique could make achievable⁴.

In Bakr and colleagues’ experiments, the final entropy (and thus temperature) was limited by heating caused by the optical-lattice light beams, as well as by inefficiencies in the excitation of atoms to higher bands. Its value was $0.27k_B$ per particle (where k_B is the Boltzmann constant), which is comparable to previously reported values¹⁰. These limitations will need to be overcome if optimal cooling is to be achieved. But even without additional cooling, the use of OEB to deterministically

control the number of atoms in individual sites will be a valuable tool for future experiments. ■

Gretchen K. Campbell is at the Joint Quantum Institute, National Institute of Standards and Technology, and the University of Maryland, Gaithersburg, Maryland 20899–8424, USA.

e-mail: gretchen.campbell@nist.gov

1. Lewenstein, M. *et al.* *Adv. Phys.* **56**, 243–379 (2007).
2. Bloch, I., Dalibard, J. & Zwerger, W. *Rev. Mod. Phys.* **80**, 885–964 (2008).
3. McKay, D. C. & DeMarco, B. *Rep. Prog. Phys.* **74**, 054401 (2011).
4. Bakr, W. S. *et al.* *Nature* **480**, 500–503 (2011).
5. Kastner, M. A. *Rev. Mod. Phys.* **64**, 849–858 (1992).
6. Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. K. *Rev. Mod. Phys.* **79**, 1217–1265 (2007).
7. Isenhower, L. *et al.* *Phys. Rev. Lett.* **104**, 010503 (2010).
8. Wilk, T. *et al.* *Phys. Rev. Lett.* **104**, 010502 (2010).
9. Baugh, J., Moussa, O., Ryan, C. A., Nayak, A. & Laflamme, R. *Nature* **438**, 470–473 (2005).
10. Bakr, W. S. *et al.* *Science* **329**, 547–550 (2010).

CONDENSED-MATTER PHYSICS

A fresh twist on shrinking materials

Unusual lattice vibrations have been discovered in scandium trifluoride — a simple compound that shrinks when heated. This finding may help to explain the phenomenon of negative thermal expansion.

J. PAUL ATTFIELD

The term might sound like a euphemism for misfortune, as in ‘negative equity’ or ‘negative patient-care outcome’, but materials that exhibit negative thermal expansion (NTE) are interesting and sometimes useful exceptions to the general rule that substances expand when heated. Negative or zero thermal expansion arises from electronic effects in some metal alloys, such as the iron–nickel alloy Invar ($\text{Fe}_{0.64}\text{Ni}_{0.36}$). Many materials in a second ‘structural’ class, in which NTE is associated with vibrations that bend an atomic network, have been discovered in the past 20 years. Deviations of the materials’ lattice vibrations from the simplest (harmonic) behaviour are expected to contribute to NTE, but most of these materials have structures that are too complex for detailed vibrational analysis. Writing in *Physical Review Letters*, Li *et al.*¹ provide an elegant demonstration of a non-harmonic (anharmonic) twisting vibration of a simple structural NTE material, scandium trifluoride (ScF_3), that helps to explain the origin of NTE in this class.

Scandium trifluoride contains a simple network of corner-sharing ScF_6 octahedra (Fig. 1). Other fluorides and oxides with the same structural arrangement do not show significant negative expansion, so it came as a surprise when a large NTE effect was discovered² in ScF_3 across a wide range of temperatures, from 10 to 1,100 kelvin. Previously reported materials with a comparable NTE, such as zirconium tungstate (ZrW_2O_8), have more complex crystal structures³, so ScF_3 has proved an ideal material in which to explore the subtle features of lattice vibrations (phonons) that can give rise to NTE.

Li and colleagues¹ used a combination of inelastic neutron-scattering experiments and phonon calculations to investigate ScF_3 . The experiments provided the phonon density-of-states, a detailed spectrum of lattice vibrations. The measurements showed that, as ScF_3 was heated, a phonon with an initial energy of about 25 millielectronvolts (meV) underwent an unusual shift to higher energies (stiffening), which is a signature of anomalous behaviour such as NTE. By comparing the measurements with the phonon calculations,

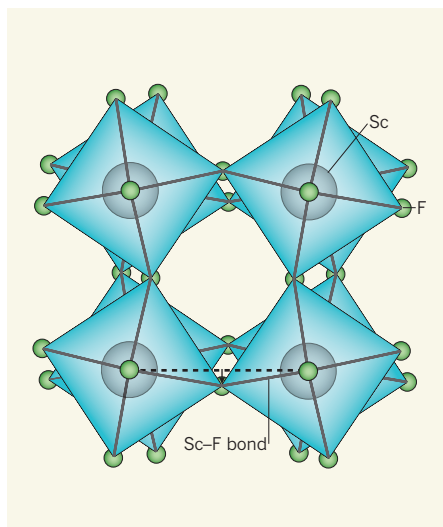


Figure 1 | Twisting vibration in the network structure of scandium trifluoride. Scandium atoms (Sc, grey) are at the centres and fluoride atoms (F, green) at the six shared corners of each of the eight ScF_6 octahedra seen here in projection. When the Sc–Sc distances are fixed (dotted line), twisting vibration stretches the Sc–F bonds. If these bonds stretch harmonically, the potential energy of the motion varies with x^4 , where x is transverse fluoride displacement (as indicated by the arrow). Li *et al.*¹ discovered that this form of vibration makes an important contribution to the phenomenon of negative thermal expansion in scandium trifluoride. (Modified from ref. 1.)

the authors demonstrated that the 25-meV vibration originates from a twisting mode of ScF_6 octahedra (Fig. 1).

A key question on structural NTE materials has been whether polyhedra such as the ScF_6 octahedra distort during twisting vibrations. If the octahedra remain rigid during the vibrations, this ‘rigid unit mode’ has a harmonic potential: the potential energy of the motion varies with the square of the amplitude (x), like the stretching of a mechanical spring. This vibration would bring the scandium atoms at the centre of the octahedra together, shrinking the lattice. However, if the Sc–Sc distances instead remain fixed so that the Sc–F bonds have to stretch during the twisting motion, the potential energy has a quartic (x^4) variation — an example of anharmonicity. Li and colleagues’ analysis¹ of ScF_3 reveals that the twisting vibration has an almost ideal quartic potential. They validate their analysis by showing that both the initial energy and the stiffening of the 25-meV phonon are consistent with the quartic potential for this twisting mode.

After modifying the phonon calculations slightly by allowing the stretched Sc–F bonds to pull the scandium atoms together, the authors demonstrate that this anharmonic model predicts NTE that agrees well with the observed thermal behaviour for ScF_3 (ref. 2).

This result shows that the non-rigidity of the ScF_6 octahedra makes an important contribution to NTE. However, this does not signal the end of the rigid-unit-mode approach: Li *et al.*¹ acknowledge that both harmonic rigid unit motions and quartic anharmonic effects are likely to be needed to provide a full explanation for NTE. With continuing improvements in inelastic neutron spectroscopy and in the computer power available for the supporting phonon calculations, the role of quartic anharmonicity in structural NTE is likely to be clarified further through future studies of more complex materials.

Finally, I note that the discovery² of NTE in ScF_3 marks another step in the seemingly unstoppable rise of perovskite-type materials. Perovskites, named after the mineral form of the calcium titanium oxide CaTiO_3 , have structures like that of ScF_3 but with an additional atom such as calcium that occupies the holes formed between groups of eight octahedra (Fig. 1). They are already

known to have excellent electronic and magnetic properties. Very large electronically driven NTE at ambient temperatures has been discovered in materials that are based on two perovskites, ZnNiMn_3 (ref. 4) and BiNiO_3 (ref. 5); now the ‘empty perovskite’ ScF_3 has been confirmed as a fundamental member of the structural NTE family. ■

J. Paul Attfield is in the Centre for Science at Extreme Conditions and the School of Chemistry, University of Edinburgh, Edinburgh EH9 3JZ, UK.
e-mail: j.p.attfield@ed.ac.uk

1. Li, C. W. *et al.* *Phys. Rev. Lett.* **107**, 195504 (2011).
2. Greve, B. K. *et al.* *J. Am. Chem. Soc.* **132**, 15496–15498 (2010).
3. Mary, T. A., Evans, J. S. O., Vogt, T. & Sleight, A. W. *Science* **272**, 90–92 (1996).
4. Takenaka, K. & Takagi, H. *Appl. Phys. Lett.* **87**, 261902 (2005).
5. Azuma, M. *et al.* *Nature Commun.* **2**, 347 (2011).

PHYSIOLOGY

On time metabolism

In mammals, molecular clocks regulate transcription and glucose homeostasis. One way they do so is by controlling glucocorticoid–receptor signalling, which suggests that clocks are embedded in liver metabolism. [SEE LETTER P.552](#)

JOSEPH BASS

The idea that genes program behaviour dates back to the 1970s, when researchers discovered that heritable mutations in fruitflies alter the daily, or circadian, sleep–wake cycle. At a molecular level, the clock machinery, which is composed of an autoregulatory transcription feedback loop¹, is present in almost all cell types and generates oscillations in the expression of at least 10% of the genes transcribed^{2,3}. Such observations led to an intensive search for the clock’s physiological functions. Lamia *et al.*⁴ describe one such function on page 552 of this issue. They report that the clock protein cryptochrome acts as an on/off switch for the nuclear receptor of glucocorticoids — hormones that are involved in glucose and lipid metabolism⁵. The finding expands our understanding of how metabolism is coordinated with geophysical time.

The impact of circadian clocks on organismal energetics and survival has been investigated in photosynthetic organisms. In cyanobacteria⁶ and plants⁷, alignment of the period length with the environmental light cycle (a defining property of the clock) influences growth and reproduction. This suggests that ‘resonance’ of internal oscillators with the environment may provide a selective

advantage during evolution. In these organisms, the clock machinery also separates oxygen- and nitrogen-fixation processes, averting accumulation of cytotoxic free radicals and preventing futile energetic cycles⁸.

In mammals, circadian rhythms operate at the behavioural, physiological and molecular levels. But studying the link between circadian cycles and metabolism in organisms with a nervous system is complicated. This is because the light response originates from the brain’s pacemaker neurons, whereas nutrient sensing occurs in other neurons and throughout many peripheral tissues. The discovery that fibroblasts maintained in culture exhibit self-sustaining circadian oscillations that show signature properties of a circadian clock⁹ was crucial because it hinted that peripheral clocks can function autonomously. Normally, however, brain and peripheral timepieces seem to communicate to give rise to coherent metabolic outputs. Indeed, genetic studies suggest that the brain and peripheral clocks participate in weight regulation and in glucose and lipid metabolism³.

Circadian oscillations in hormone levels are certainly crucial for metabolic health, and disturbances in these have been used in clinical diagnosis for decades. For instance, high blood levels of the glucocorticoid hormone cortisol

at midnight are regarded as a sign of Cushing's disease. Similarly, a 'dawn phenomenon' of elevated blood glucose levels arises in humans with diabetes owing to increased production of glucocorticoids in the early morning.

At a molecular level, expression of the nuclear receptors for glucocorticoids shows strong circadian oscillations¹⁰. In fact, certain nuclear hormone receptors are integral components of the clock machinery¹¹. These receptors' activity results in a complex pattern of gene activation and repression, affecting metabolic and haematopoietic tissues in various ways. The clock protein PER2 interacts with several nuclear hormone receptors¹², but the physiological impact of the clock machinery on the receptors' function has not been established.

Lamia and colleagues previously showed¹³ that the clock-repressor protein cryptochrome (Cry) is a sensor of cellular energy state. To further dissect the coupling between Cry and metabolic signals, some of the same authors now survey interactions of Cry with a broad range of nuclear hormone receptors in mice⁴.

The authors report a strong and specific interaction between Cry and glucocorticoid receptors that shifts the balance towards gene repression after activation of these receptors. In addition, Cry interacted with two nuclear hormone receptors that activate transcription of the clock-activator protein Bmal1. This suggests that Cry may contribute to the oscillator's robustness by participating in a short feedback loop controlling Bmal1 expression. Intriguingly, Cry interaction with glucocorticoid receptors was enhanced by dexamethasone⁴ — a drug that acts as a glucocorticoid. Because circadian levels of Cry vary, this observation directly links glucocorticoid signalling to circadian-clock oscillations (Fig. 1). Lamia and collaborators also note that Cry modulates patterns of gene activation and repression stimulated by glucocorticoid receptors. In Cry-deficient cells, glucocorticoid signalling led to a higher number of activated targets, and a vast decrease in the number of genes normally repressed by this clock protein.

Because the main outcome of glucocorticoid-receptor activity is glucose generation in the liver by the process of gluconeogenesis, and because Cry-deficient animals display impaired glucose tolerance¹⁴, Lamia *et al.*⁴ focused on Cry function in the liver. The authors found that Cry directly represses the expression of *Pck1* — a gene that encodes the rate-limiting enzyme in gluconeogenesis (Fig. 1). In Cry-deficient cells, dexamethasone increased *Pck1* expression, but did not affect repression of genes mediating the inflammatory NF- κ B signalling pathway. This observation indicates a separation of Cry functions in the gluconeogenic and inflammatory pathways.

Cry deficiency also increased the sensitivity

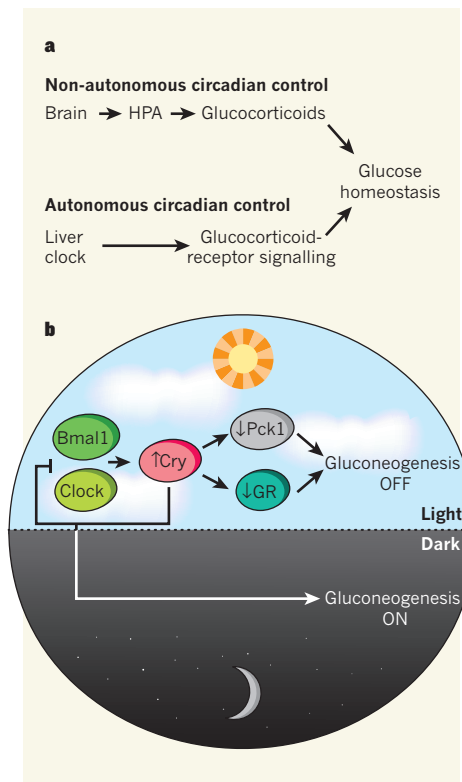


Figure 1 | Biological clocks organize glucose metabolism. **a**, Glucose homeostasis is regulated by the interplay between non-autonomous circadian control — which affects glucocorticoid hormone levels through the brain and the hypothalamic–pituitary–adrenal (HPA) axis — and autonomous oscillatory regulation of these hormones' signalling in peripheral tissues such as the liver. Lamia *et al.*⁴ report that the clock-repressor protein Cry plays a central part at both of these levels. **b**, In the liver, the clock-activator proteins Clock and Bmal1 increase Cry transcription, which downregulates gluconeogenesis both by interacting with — and so inhibiting — glucocorticoid receptors (GRs), and by reducing the expression of the *Pck1* enzyme. As is typical of the clock machinery, high levels of Cry also downregulate Clock and Bmal1 transcription in a negative feedback loop, thus reducing its own levels and so promoting gluconeogenesis during the dark phase of the light–dark cycle.

of mice to glucocorticoid-induced diabetes⁴. This finding has therapeutic implications given the widespread use of steroids — of which glucocorticoids are an example — for the management of inflammation associated with cancer and conditions related to the immune system. Increasing Cry levels may enhance the anti-inflammatory effects of glucocorticoids by blocking the adverse effects of these hormones on glucose metabolism.

Whereas previous work¹⁴ implicated Cry in regulating the signalling of glucagon (another hormone that mediates glucose homeostasis) in the cytoplasm, this paper pinpoints another role for Cry in the control of gluconeogenesis at the nuclear level. The modulation of

glucocorticoid signalling by Cry to control glucose homeostasis may be relevant for understanding glucose overproduction by the liver in human diabetes, not least because a genomic-association study¹⁵ indicated that *CRY2* gene variants are associated with glucose homeostasis in humans.

At the organismal level, glucocorticoid regulation in animals is controlled through a classical neuroendocrine feedback loop involving communication between the hypothalamus, pituitary and adrenal tissues (Fig. 1). Lamia *et al.* find that, in Cry-deficient mice, blood glucocorticoid levels remain high at all times of the day–night cycle, with the hormone's overproduction arising at the level of the hypothalamus. How Cry deficiency might impair feedback inhibition of steroidogenesis in the hypothalamus is a fascinating question.

Lamia and colleagues' findings may also be relevant to the spike seen in metabolic disorders among individuals subjected to shift work, sleep curtailment and other forms of circadian disruption. A feature of these disorders may be misalignment of oscillations in blood levels of glucocorticoids with oscillations in both the interaction between Cry and glucocorticoid receptors, and their signalling in the liver. This study provides an additional example of the extensive molecular integration and relatedness of temporal behavioural programmes with cell metabolism. ■

Joseph Bass is in the Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. He is also at the Center for Sleep and Circadian Biology, Department of Neurobiology, Weinberg College of Arts and Sciences, Northwestern University.

e-mail: j-bass@northwestern.edu

1. Rosbash, M. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **72**, 75–83 (2007).
2. Panda, S. *et al.* *Cell* **109**, 307–320 (2002).
3. Ramsey, K. M. & Bass, J. *Cold Spring Harb. Symp. Quant. Biol.* <http://dx.doi.org/10.1101/sqb.2011.76.010546> (2011).
4. Lamia, K. A. *et al.* *Nature* **480**, 552–556 (2011).
5. Hollenberg, S. M. *et al.* *Nature* **318**, 635–641 (1985).
6. Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S. & Johnson, C. H. *Proc. Natl Acad. Sci. USA* **95**, 8660–8664 (1998).
7. Dodd, A. N. *et al.* *Science* **309**, 630–633 (2005).
8. Rutter, J., Reick, M. & McKnight, S. L. *Annu. Rev. Biochem.* **71**, 307–331 (2002).
9. Balsalobre, A., Damiola, F. & Schibler, U. *Cell* **93**, 929–937 (1998).
10. Yang, X. *et al.* *Cell* **126**, 801–810 (2006).
11. Alenghat, T. *et al.* *Nature* **456**, 997–1000 (2008).
12. Schmutz, I., Ripperger, J. A., Baeriswyl-Aebischer, S. & Albrecht, U. *Genes Dev.* **24**, 345–357 (2010).
13. Lamia, K. A. *et al.* *Science* **326**, 437–440 (2009).
14. Zhang, E. E. *et al.* *Nature Med.* **16**, 1152–1156 (2010).
15. Dupuis, J. *et al.* *Nature Genet.* **42**, 105–116 (2010).

2011

EDITORS' CHOICE

Adapted extracts from
selected News & Views
articles published this year.

EPIDEMIOLOGY

HOW COMMON IS AUTISM?

Catherine Lord (*Nature* 474, 166–168; 2011)

Autism spectrum disorders (ASDs) encompass autism, Asperger's syndrome and other related conditions. Core features include difficulties in basic social and communicative behaviours, as well as repetitive behaviours and restricted interests. The prevalence of ASDs has been considered to be just over 1%. However, in a provocative, carefully executed study, Kim *et al.* present evidence for surprisingly high rates of these disorders — 3.74% in males and 1.47% in females — in school-age children in a South Korean community. The study is remarkable in its attempt to identify ASDs in children in mainstream schools, as well as in children receiving special care. Yet a larger question arises. Should individuals who are without impairment or disability — that is, without any suffering, limitations or restrictions in daily functioning — be diagnosed with an ASD?

Am. J. Psychiatry 168, 904–912 (2011).

ORGANIC CHEMISTRY

OVERCOMING CATALYTIC BIAS

Daesung Lee (*Nature* 471, 452–453; 2011)

The alkene metathesis reaction, in which carbon–carbon double bonds are redistributed between alkene molecules, has had an enormous impact on chemical research and industry, as was recognized in 2005 with a Nobel prize. But this remarkable reaction has an inherent limitation — it cannot generate the thermodynamically less stable isomers of the alkene products. Hoveyda and colleagues now report that they have overcome this problem using specially designed molybdenum catalysts. Several aspects of these new metathesis reactions remain to be further refined: the range of alkene substrates that can be used should be broadened beyond enol ethers and allylic amides, for example, and ways should be found to avoid using an excess of one of the reagents. Nevertheless, these reactions are highly promising and will potentially be of use for the preparation of numerous compounds, with far-reaching consequences for the future of metathesis chemistry.

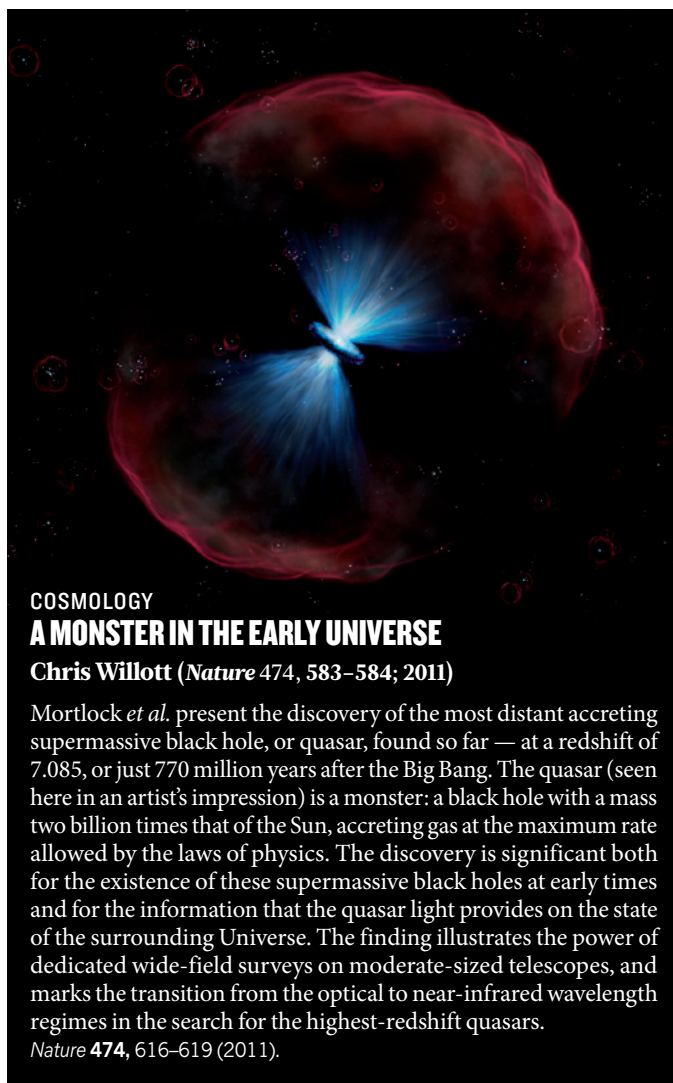
Nature 471, 461–466 (2011).

COMPLEX SYSTEMS

UNZIPPING ZIPF'S LAW

Lada Adamic (*Nature* 474, 164–165; 2011)

Perhaps the only thing more abundant in both natural and man-made systems than power laws are the models that have been developed to explain them. Baek *et al.* now argue that because such models depend on the specifics of each system, they fail to capture the shared cause of this regularity. The authors instead propose a general model that can be applied to any division of items into groups, and that can, for example, account for Zipf's law of word frequencies in text, the popularity



GEMINI OBSERV.

COSMOLOGY

A MONSTER IN THE EARLY UNIVERSE

Chris Willott (*Nature* 474, 583–584; 2011)

Mortlock *et al.* present the discovery of the most distant accreting supermassive black hole, or quasar, found so far — at a redshift of 7.085, or just 770 million years after the Big Bang. The quasar (seen here in an artist's impression) is a monster: a black hole with a mass two billion times that of the Sun, accreting gas at the maximum rate allowed by the laws of physics. The discovery is significant both for the existence of these supermassive black holes at early times and for the information that the quasar light provides on the state of the surrounding Universe. The finding illustrates the power of dedicated wide-field surveys on moderate-sized telescopes, and marks the transition from the optical to near-infrared wavelength regimes in the search for the highest-redshift quasars.

Nature 474, 616–619 (2011).

of last names, and city and county populations. A model purporting to explain a power-law distribution should be as general as Baek and colleagues' model, or it should be able to reproduce additional features of the system it models, beyond the familiar straight line on a log–log plot.

New J. Phys. 13, 043004 (2011).

PALAEOANTHROPOLOGY

MALAPA AND THE GENUS *HOMO*

Fred Spoor (*Nature* 478, 44–45; 2011)

Following on from the announcement last year by Berger *et al.* of the remains of a newly discovered hominin species, *Australopithecus sediba*, the same group has now published five reports detailing additional fossils and further analyses. Cave deposits at the Malapa site in South Africa yielded two partial skeletons, which Pickering *et al.* have found to be 1.977 ± 0.002 million years old. These skeletons are not only well preserved and remarkably complete, but also show a



STEM CELLS

THE DARK SIDE OF INDUCED PLURIPOTENCY

Martin F. Pera (*Nature* 471, 46–47; 2011)

Induced pluripotent stem cells (iPSCs) are generated through the reprogramming of differentiated adult cells and can be coaxed to develop into a wide range of cell types. But the ultimate value of these cells for research and regenerative medicine will depend on the fidelity of their reprogramming to the pluripotent state, and on their maintenance of a normal genetic and epigenetic (involving aspects other than DNA sequence) status. Five recent surveys show that the reprogramming process and subsequent culture of iPSCs *in vitro* can induce genetic and epigenetic abnormalities in these cells. The studies raise concerns over the implications of such aberrations for future applications of iPSCs. They provide little insight into the crucial question of what aspects of the reprogramming methods might predispose the cells to the accumulation of recurrent genetic or epigenetic lesions. Also it is unclear how best to assess the effects of new genetic lesions on the growth, differentiation, tumorigenicity and functionality of pluripotent stem cells or their differentiated progeny.

Nature 471, 58–62, 63–67, 68–73 (2011); *Cell Stem Cell* 7, 521–531 (2010); *Cell Stem Cell* 8, 106–118 (2011).

FORUM Ageing
LONGEVITY HITS A ROADBLOCK

(Nature 477, 410–411; 2011)

Increased expression of sirtuin proteins has been shown to enhance lifespan in several organisms. New data indicate that some of the reported effects may have been due to confounding factors in experimental design.

A VALUABLE BACKGROUND CHECK

David B. Lombard & Scott D. Pletcher

Burnett and colleagues' in-depth study of the influence of Sir2 overexpression on lifespan in *Caenorhabditis elegans* and *Drosophila*, as well as on diet-modulated longevity in the fly, shows that the protein has no effect. Given the demonstrated importance of sirtuins in mammals, why rehash the precise role of Sir2 in worm and fly ageing? First, challenging published results is an essential, self-correcting aspect of science. Second, invertebrate models continue to contribute to the understanding of sirtuin biology in mammals, and so future studies must be interpreted in the context of these new data. Third, the new reports reinforce the importance of rigorous genetic background control when interpreting the effects of single gene mutations.

DON'T WRITE SIRTUINS OFF

Carles Cantó & Johan Auwerx

Although in light of Burnett and colleagues' findings the appeal of sirtuins as a *sensu stricto* lifespan determinant might be gone, SIRT1 activation remains a promising approach to delaying general age-related physiological decline and to treating numerous inherited and acquired diseases characterized by defective mitochondrial function. The astonishing ability of SIRT1 to enhance 'healthspan' by promoting metabolic fitness will also guarantee it a long life as a subject for further exciting research.

Nature 477, 482–485 (2011).

surprising mix of morphological characters. The authors conclude that *A. sediba* fossils as late as those preserved at Malapa could have been the ancestor of *Homo*. It will, however, be difficult to uphold the suggestion that the extensive evolutionary change required could have occurred in the time available (a maximum of 80,000 years) if *A. sediba* at Malapa gave rise to *Homo* species.

Science 333, 1402–1407, 1407–1411, 1411–1417, 1417–1420, 1421–1423 (2011).

PRECISION MEASUREMENT

A SEARCH FOR ELECTRONS THAT DO THE TWIST

Aaron E. Leanhardt (*Nature* 473, 459–460; 2011)

If I were to tell you about an elementary particle that has mass and charge, but neither size nor structure, yet still has a well-defined orientation and can point in a specific direction in space, you would probably think I am describing something from a science-fiction novel. In fact, I am telling you about the electron. Hudson *et al.* describe an experiment aimed at refining our understanding of this fundamental particle and, more broadly, the basic laws of nature. Described colloquially, their experiment searches for evidence of an aspheric distortion to the shape of the electron. Hudson *et al.* observe no such distortion. The authors have pioneered the use of cold polar molecules to push the search for an electric dipole moment of the electron to new levels. Establishing the existence of an electric dipole moment of a fundamental particle is an exclusively experimental endeavour. Hudson *et al.* are the latest to attempt such a feat. Experiments of this genre reach far beyond the realm of atomic, molecular and optical physics: they can be viewed as low-energy windows on the high-energy soul of the cosmos.

Nature 473, 493–496 (2011).

BEHAVIOURAL NEUROSCIENCE

FRUITY APHRODISIACS

Benjamin Prud'homme & Nicolas Gompel
(*Nature* 478, 190–191; 2011)

The smell of a delicious stew often stimulates a man's appetite, but it rarely turns him on. Male *Drosophila* fruitflies, however, behave differently. Grosjean *et al.* identify how particular odours from rotten fruit and decaying vegetables — the creatures' typical diet — act as aphrodisiacs to male fruitflies. The authors show that the neural circuit mediating male-fly courtship is equipped with a plug-in to sense food resources that modulates the decision to court females. Regardless of how male flies integrate food odours into the decision to mate, one may wonder why they do so in the first place. Linking courtship behaviour to

a food compound gives the male some sway over how his progeny will be raised. Should the levels of phenylacetic acid be limited (read: low food resources), a male might not insist on courting local females, no matter how attractive they smell, and might wait to sire progeny grown in a richer environment.

Nature 478, 236–240 (2011).

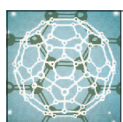


Rethinking amide bond synthesis

Vijaya R. Pattabiraman¹ & Jeffrey W. Bode¹

One of the most important reactions in organic chemistry—amide bond formation—is often overlooked as a contemporary challenge because of the widespread occurrence of amides in modern pharmaceuticals and biologically active compounds. But existing methods are reaching their inherent limits, and concerns about their waste and expense are becoming sharper. Novel chemical approaches to amide formation are therefore being developed. Here we review and summarize a new generation of amide-forming reactions that may contribute to solving these problems. We also consider their potential application to current synthetic challenges, including the development of catalytic amide formation, the synthesis of therapeutic peptides and the preparation of modified peptides and proteins.

Amide linkages¹ are not only the key chemical connections of proteins but they are also the basis for some of the most versatile and widely used synthetic polymers. Chemical reactions for their formation are among the most executed transformations in organic chemistry (Fig. 1). The prevalence of amide functionality, particularly in peptides and proteins², sometimes gives the incorrect impression that there are no remaining synthetic challenges. This is surprising, as it is often the case that even simple amides resist formation, forcing practitioners to resort to ever more exotic and expensive reagents for their synthesis. Furthermore, the favourable properties of amides, such as high polarity, stability and conformational diversity, make it one of the most popular and reliable functional groups in all branches of organic chemistry. Improved methods for the synthesis of amide functionality, whether catalytic and waste-free or chemoselective and suitable for fragment coupling, are in great demand.



2011: YEAR OF CHEMISTRY
Celebrating the central science
nature.com/chemistry2011

In living systems, most amide bonds are formed by the complex factories that are ribosomes. Long, complex proteins are assembled amino acid by amino acid, using a templated amidation of amines and the active esters of amino acid monomers and RNA (Fig. 2a)³. Synthetic chemists, by contrast, do not have the luxury of working on this single-molecule scale, and instead deal with trillions of molecules that must be coaxed into precise reaction trajectories. This strategy necessitates that nearly every functional group be protected by a bulky hydrophobic appendage, leading to a reliable, but rather wasteful approach to peptide synthesis, in which dozens of molecules are sacrificed to form just one amide bond⁴.

The current methods for amide formation are remarkably general but at the same time widely regarded as expensive and inelegant. Not surprisingly, in 2007 the American Chemical Society Green Chemistry Institute (comprising members from major pharmaceutical industries worldwide) voted 'amide formation avoiding poor atom economy reagents' as the top challenge for organic chemistry⁵. Furthermore, even the best stoichiometric reagents often fail for the synthesis of sterically hindered amides. The issues of waste and expense associated with amide formation are compounded when applied to peptide synthesis, and are responsible for the great cost of commercial therapeutic peptides. The chemical synthesis of proteins is largely prohibited by limitations inherent to traditional amide formation, although advances in the chemoselective ligation of unprotected peptide fragments demonstrate how advances in amide-forming methodologies can have far-reaching impacts across scientific disciplines.

The state of the art in amide bond formation

Before considering emerging methodologies for amide bond construction, we should take stock of the successes of existing amide-forming reactions. Acylation of amines with activated carboxylic acids is the most common reaction performed in the synthesis of modern pharmaceuticals, accounting for 16% of all reactions, according to a recently analysed data set of medicinal chemistry campaigns⁶. In the peptide field, the development of solid-phase peptide synthesis⁷ and subsequent improvements in coupling reagents, protecting groups, resins and chromatographic methods have made the synthesis of small amounts (5–1,000 mg) of moderately sized (30–50 residues) peptides routine and commonplace (Fig. 2b). Hydrophobic peptide sequences and challenges with non-canonical amino acids still create problems, but there are few peptides of this size that cannot be produced by a skilled practitioner on a laboratory scale. For the synthesis or semi-synthesis of proteins,

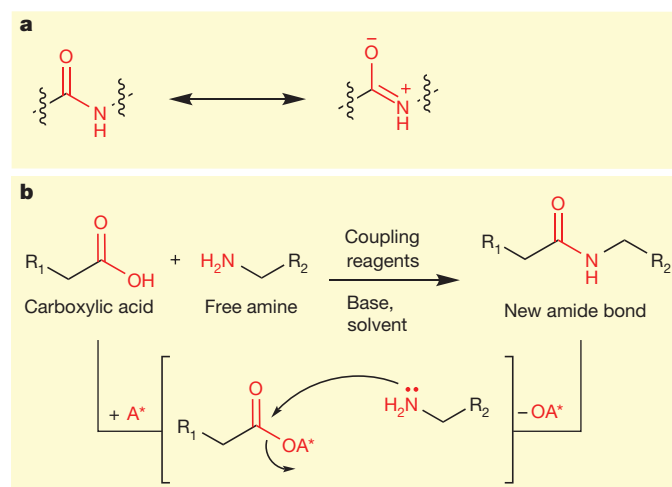


Figure 1 | Chemical structure of amides and the conventional chemical method for amide bond synthesis. **a**, Resonance structures of an amide group. **b**, The conventional method for formation of an amide bond. It involves activation of a carboxylic acid by an activating group (A*), followed by nucleophilic displacement by a free amine to generate a new amide bond in the presence of coupling reagent, base and solvent. R₁ and R₂, small molecules, peptides or proteins.

¹Laboratorium für Organische Chemie, ETH Zürich, 8093 Zürich, Switzerland.

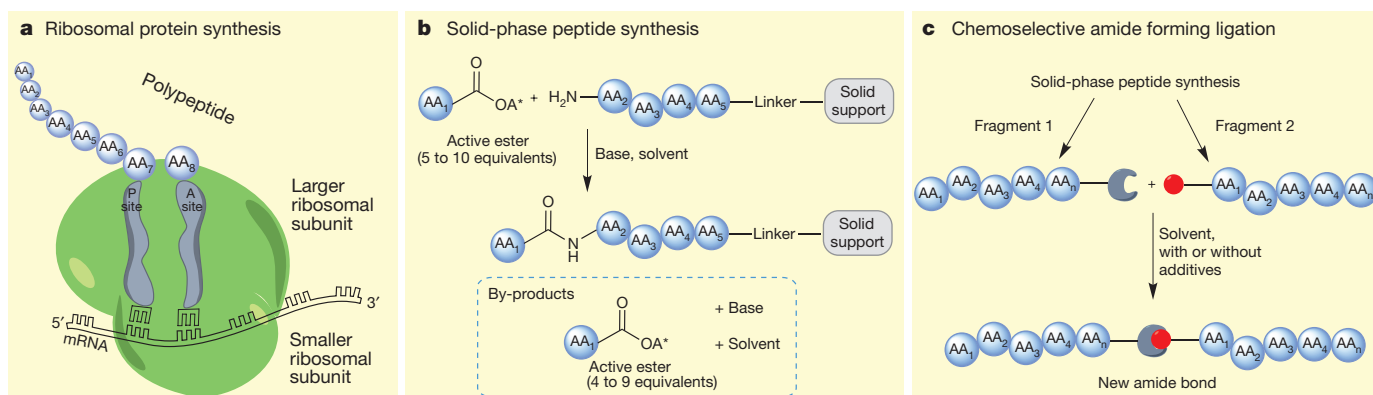


Figure 2 | Representation of protein and peptide synthesis by biochemical and chemical methods. **a**, Simplified cartoon depiction of protein synthesis by ribosomes. Proteins are assembled by templated amidation with unprotected amino acids (blue spheres, AA_x). The A site (grey, right) serves as the entry point for aminoacyl transfer RNA and elongation of the polypeptide occurs along the P site (grey, left) by formation of a new amide linkage. This process is highly efficient and readily produces very high-molecular-weight proteins. mRNA, messenger RNA. **b**, Schematic illustration of routinely used solid-phase peptide synthesis. Peptides are assembled on a solid support with active esters

generated from fully protected amino acid monomers, coupling reagent and base. Peptides between 30 and 50 residues are routinely produced with natural or unnatural amino acids. **c**, Scheme depicting the concept of chemoselective amide forming ligation for the synthesis of complex peptides, proteins and glycopeptides. Two uniquely reactive functional groups (represented by filled red circle and matching grey shape) on unprotected fragments react chemoselectively to readily produce proteins up to 200 amino acid residues with fewer by-products and waste.

native chemical ligation⁸ (NCL) has revolutionized the field and made possible access to materials once thought impossible (Fig. 2c). Examples include the preparation of enantiomeric proteins such as snow flea antifreeze protein⁹, plectasin¹⁰ and the 203 amino acid HIV-1 protease covalent dimer¹¹. An often overlooked achievement in synthetic amide chemistry is the access to high-molecular-weight amide-based polymers, including nylons and aramids, made possible by advances in polymerization techniques, among others¹².

Emerging chemical methods for amide bond formation

The limits of traditional amide-forming reactions and NCL will continue to be pushed and tested, but the next breakthrough may come from new, unexpected and selective methods for amide bond formation. Although many of the emerging methods described below are still in their infancy, they reflect a growing creativity and urgency among synthetic organic chemists in addressing what is increasingly recognized as an unmet synthetic need. This emerging generation of new amide formations may provide the basis for important catalytic methods or novel fragment coupling strategies.

The conventional approach to amide formation is the condensation of an amine with a carboxylic acid via an active ester. New methods for amide formation fall into two broad categories classified by their reaction partners. In the first, amines serve as one of the reaction partners and are acylated by either a catalytic or oxidative fashion. In the second, novel combination of reaction partners lead to amide products by mechanistically unique pathways.

New acylation reactions of amines

Amidation by catalytic acylation of amines with carboxylic acids. Traditional amide synthesis relies on a coupling reagent to convert an unreactive carboxylic acid into an activated carboxylate for reaction with amines to give amides. In the absence of a coupling reagent, the carboxylic acid and the amine simply form a carboxylate-ammonium salt, rather than an amide product, owing to the unfavourable thermodynamics of the amide-forming reaction. Despite this, several promising catalysts for direct amide synthesis from carboxylic acid and amine in the absence of a coupling reagent have begun to appear^{13,14}. Most prominent are the boronic acid catalysts, first reported in ref. 15, and with recent improvements^{16–18} (Fig. 3a). In all these cases, the boronic acid takes the role of a coupling reagent in generating an active ester suitable for amidation in a waste-free catalytic manner.

Amidation by catalytic generation of activated carboxylates. A growing and important concept in organic chemistry is the concept of redox economy, in which internal exchange of oxidation states between adjacent functional groups within or between reactants provides access to a reactive intermediate without the need for stoichiometric reagents. In the context of amide bond formation, this has been achieved by the catalytic generation of activated carboxylates from functionalized aldehydes, such as formylcyclopropanes, α,β -unsaturated aldehydes¹⁹ and their more conveniently prepared α -hydroxyketone surrogates²⁰, α -haloaldehydes and epoxyaldehydes²¹. These processes work particularly well with a *N*-heterocyclic carbene (NHC) catalyst and a co-catalyst to generate an activated carboxylate, which is then converted to an amide with a variety of amines (Fig. 3b). Such reactions make possible amide formation using only catalytic amounts of reagents, and offer a promising solution to the recognized problem of side-product generation in traditional amide formations. The mechanistic basis for these catalytic reactions requires more elaborate starting materials than classical acylating agents generated from carboxylic acids. When used to access targets for which the classical monomers are not readily available, such as β -amino acids, the initial synthetic burden for the preparation of these functionalized monomers is acceptable, particularly if the subsequent amide formations do not require stoichiometric coupling reagents. Enantioselective variants of new NHC-catalysed reactions for amidation with functionalized aldehyde monomers are an area of intense interest, suggesting that these reactions may one day lead to catalytic, enantioselective peptide synthesis.

Catalytic, oxidative amidation of amines. An alternative redox approach to amide formation is acylation of amines from alcohols or aldehydes along with a stoichiometric oxidant. NHC catalysts have been used for amide formation from simple aldehydes, stoichiometric oxidants and a suitable co-catalyst (Fig. 3b)²². Several metal catalysts and oxidants^{23,24} have also been identified for this process, with the most promising being palladium²⁵ and copper/silver²⁶ (Fig. 4a).

An intriguing example of oxidative amidation from the alcohol oxidation state is catalytic ruthenium promoted coupling of amines and alcohols²⁷. In general, this reaction proceeds in a clean atom-economical manner without any acid, base or additives, and generates molecular hydrogen as the only by-product (Fig. 3c). Thus, although it is formally an oxidative amidation, no stoichiometric oxidant is needed. Mechanistically, the ruthenium pincer-type catalyst promotes dehydrogenation of the alcohol to aldehyde, which then participates in amidation with amine

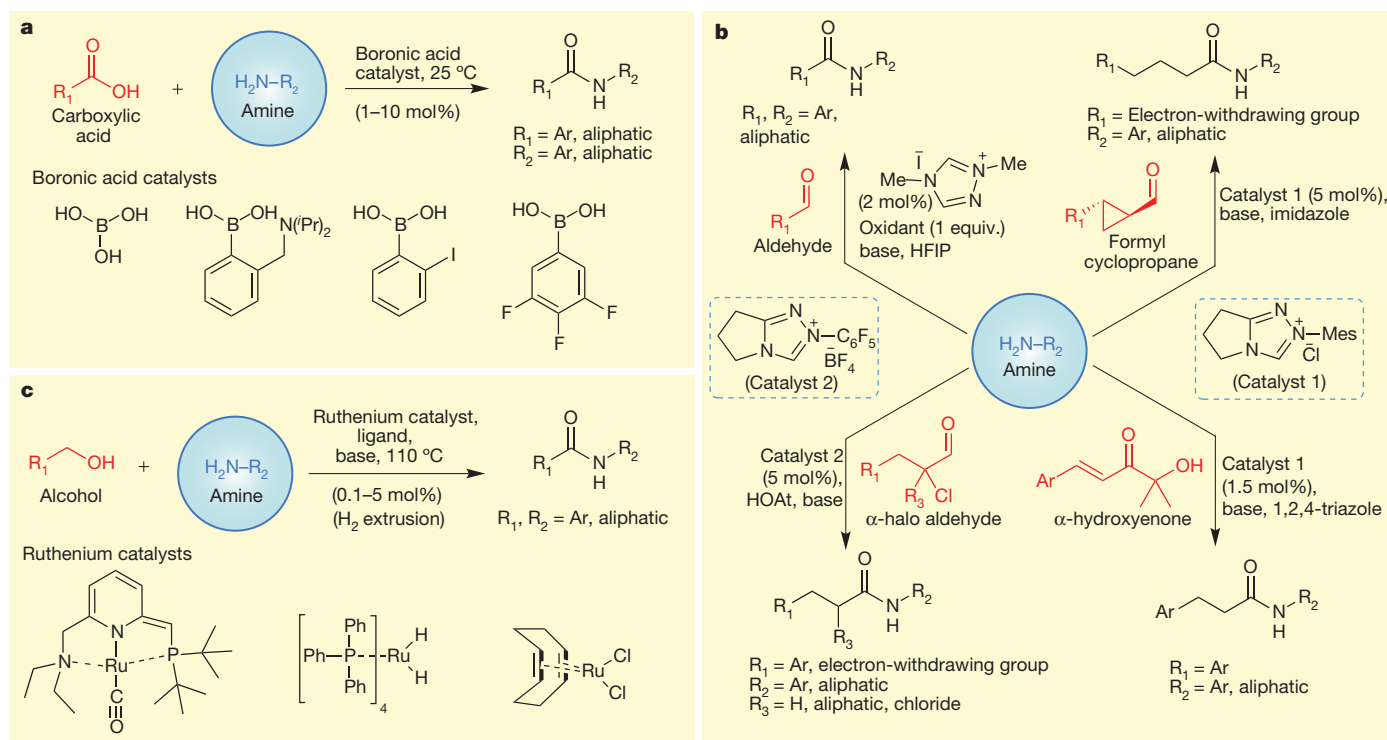


Figure 3 | Emerging organocatalytic and metal-catalysed methods for amide synthesis. **a**, Catalytic direct condensation (top) of carboxylic acid (red) and amine (blue) in the presence of many promising boronic acid catalysts (bottom) for the synthesis of aliphatic and aromatic amides. **b**, Organocatalytic redox and oxidative amidation strategies for chemoselective amide formation by reaction of amines (blue, middle) with aldehydes, formylcyclopropanes,

α -hydroxyketones and α -haloaldehydes (red). **c**, Direct coupling (top) of an alcohol (red) with an amine (blue) to obtain amides by hydrogen extrusion in the presence of ruthenium catalysts (bottom). Ar, aromatic; Pr , isopropyl; Me, methyl; HFIP, hexafluoroisopropanol; Mes, 2,4,6-tri-methylphenyl; HOAt, 3H-1,2,3-triazolo[4,5-b]pyridin-3-ol; Ph, phenyl.

guided by the catalyst. Several alternative ruthenium catalysts have been identified for this transformation, with limited improvement over the original ruthenium pincer catalyst^{28,29}.

Other than carboxylic acids, aldehydes and alcohols, alkynes have also been used as acyl surrogates in reactions with amines to give amides. Aromatic and aliphatic alkynes have been shown to react in aqueous basic-oxidative conditions under the control of a manganese-porphyrin catalyst to generate a postulated ketene intermediate that reacts selectively to give an amide (Fig. 4a)³⁰. In unprotected peptides, the amidation chemoselectively occurs only at the amino (N)-terminal α -amino group in the presence of side-chain amine, hydroxyl and amide groups.

Oxidative coupling of α -bromo nitroalkanes with amines. A quite different approach to amide formation is the oxidative coupling of α -bromo nitroalkanes with amines in the presence of *N*-iodosuccinimide and potassium carbonate, reported in 2010 (Fig. 4b)³¹. It is proposed that the α -bromo nitroalkane assumes the role of a nucleophile and that the *N*-iodo amine component serves as the electrophile, essentially leading to a reversal of traditional roles of an acyl donor and amine. This method

is particularly attractive for the preparation of arylglycine-containing peptides, for which the necessary amino acid monomers are difficult to obtain and the peptide couplings prone to epimerization.

Amidation by acylation with amine surrogates

Just as chemists have developed amidations from alternatives to carboxylic acid starting materials, efforts to use surrogates for the traditional amine reaction partner have intensified in recent years. Towards this end, isonitriles have been shown to react with carboxylic acids³² to generate a formimidate carboxylate mixed anhydride. This intermediate rearranges to a *N*-formyl amide under high temperature microwave conditions, which can then be converted to a primary amide by removal of the *N*-formyl group with a base. The harsh reaction conditions and the restricted functional group tolerance to some extent limits the application of this reaction for sensitive peptides or biomolecules. Nevertheless, very recently, cyclosporine A, a highly *N*-methylated cyclic peptide, has been synthesized in good yields by isonitrile-mediated coupling with peptide carboxylic acids or thioacids (Fig. 5a)³³. In general,

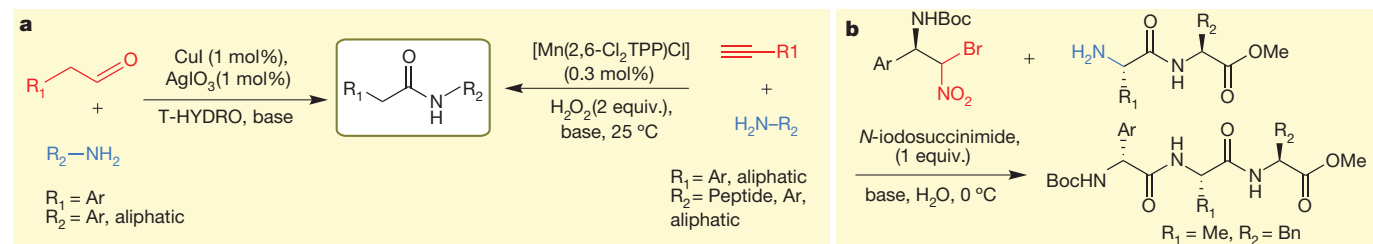


Figure 4 | Metal-catalysed and oxidative methods for amide synthesis. **a**, Left: copper-catalysed direct formation of amides from amines (blue) and aldehydes (red). Right: a manganese-porphyrin-catalysed amidation reaction between alkynes (red) and amines (blue) involving a stoichiometric oxidant.

b, Oxidative coupling of α -bromo-nitro compounds (top left) with *N*-iodoamines (intermediates, not shown) for the synthesis of peptides (bottom right). Boc, *N*-t-butoxycarbonyl; Bn, benzyl; T-HYDRO, *t*-butyl hydroperoxide; TPP, meso-tetraphenylporphine.

for the synthesis of difficult *N*-methylated peptides, phosphonium-based coupling reagents and bis(trichloromethyl)carbonate have been the methods of choice and now isonitrile chemistry presents interesting possibilities³⁴.

Several new chemoselective approaches to amides from thioacids have recently appeared (Fig. 5b), based on pioneering, but long-overlooked chemistry^{35–38}. These include the coupling of thioacids with azides³⁹ to give amides without the use of activating or coupling agents. Mechanistically, the reaction between thioacid and azide is believed to generate a thiatriazoline intermediate that breaks down to an amide, liberating nitrogen and elemental sulphur. Thioacids, in contrast to carboxylic acids, have been shown^{40,41} to react with isonitriles at ambient temperature to produce a thioformimidate carboxylate mixed anhydride intermediate, which could then be reacted with a peptide amine to give an amide⁴². In related reactions, thioacids have been reacted with electron-deficient sulphonamides^{43,44}, isocyanates and isothiocyanates to produce amide linkages⁴⁵. These reactions can even be performed in the presence of unprotected alcohols, partially protected peptides and highly hindered amino acids, and may hold promise for the challenging area of glycopeptide synthesis.

Transamidation reactions

An interesting reaction in the context of amide chemistry is an amide-metathesis or transamidation process that would exchange the constituents of two different amide groups. Efforts towards this goal have been reported⁴⁶. Although secondary and tertiary amides are extremely inert under normal conditions, they can be activated in the presence of aluminium, zirconium or hafnium-amido catalysts to undergo transamidation and amide metathesis reactions. These reactions could

allow secondary and tertiary amides to be used as substrates in dynamic covalent chemistry and oligoamide synthesis. Although the reported conditions are not yet suitable for reactions on proteins or unprotected peptides, one could imagine applications such as chemoselective chemical protein splicing to excise a protein segment of interest from a full-length protein for further studies and manipulations. As more mild and chemoselective metal-catalysed or enzymatic transamidation reactions become available, this method could become a valuable tool in protein engineering or in the preparation of bio-inspired materials.

New methods for chemoselective ligation

One of the greatest needs in new methods for amide formation is the discovery of chemoselective amide-forming ligations, ideally those that can lead to amide products in the presence of unprotected amines, carboxylic acids and alcohols. NCL⁸, in which a carboxy (C)-terminal peptide or protein thioester reacts with an N-terminal cysteine peptide or protein to give an amide-linkage, represents a nearly ideal chemoselective amide formation, in that it operates under mild aqueous conditions without complications from unprotected side chains (Fig. 6a). Its limitation is the need for an N-terminal cysteine residue and occasional difficulties in preparing the C-terminal thioesters, particularly by Fmoc solid-phase peptide synthesis. To circumvent these problems, approaches have been developed for NCL with cysteine-like amino acid residues⁴⁷ and for accessing C-terminal thioesters^{48,49}. The wide utility and reliability of NCL has elevated it to a privileged place among chemical methods for peptide, protein and glycopeptide synthesis. Inspired by these advances, chemists have sought to identify new amide-ligation strategies that mirror the functional group tolerance and rapid reaction rates of NCL. Given the rare occurrence of cysteine in

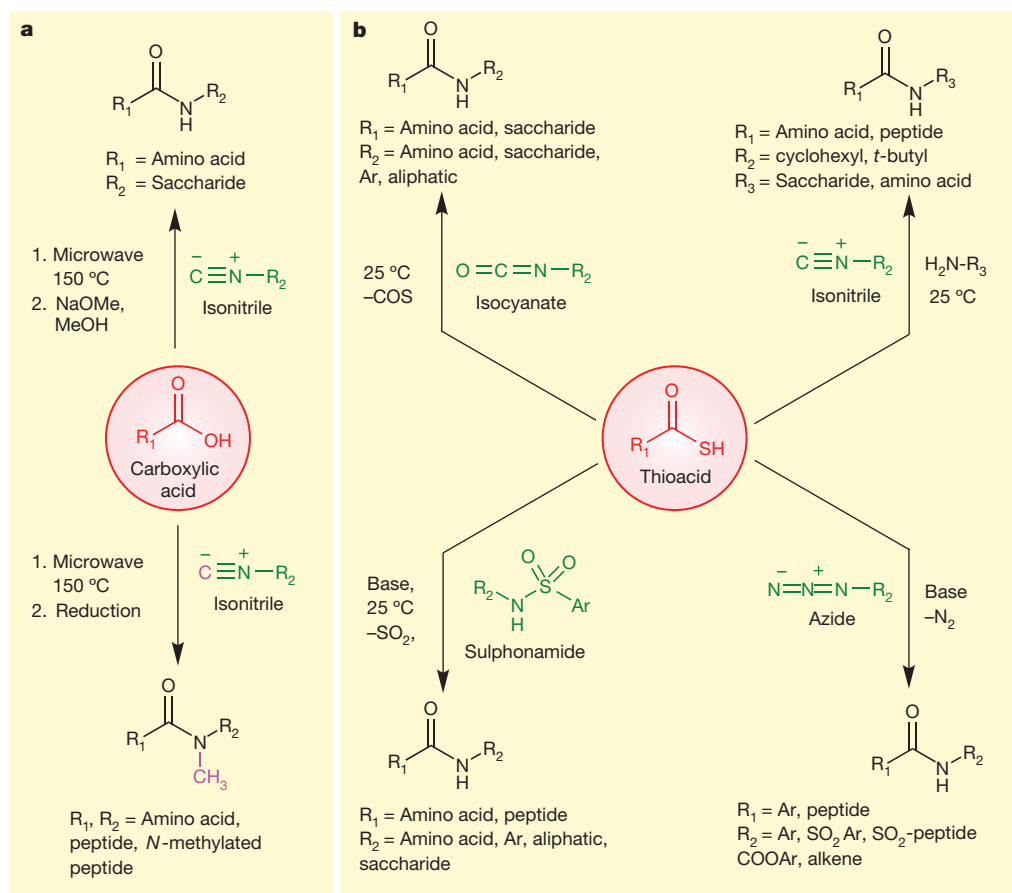


Figure 5 | Emerging reactions for chemoselective amide bond formation with carboxylic acid, thioacid and amine surrogates. **a**, Direct condensation of carboxylic acid (red, middle) with isonitriles under microwave conditions for synthesis of amino acids, *N*-methylated peptides and glycopeptides. **b**, Reaction

of a thioacid (red, middle) with various amine surrogates, such as an azide, an electron-deficient sulphonamide, isocyanate and isonitrile, to give, in many cases, a native amide linkage containing amino acid, peptide, saccharide, aromatic and aliphatic compounds. *t*-, tertiary.

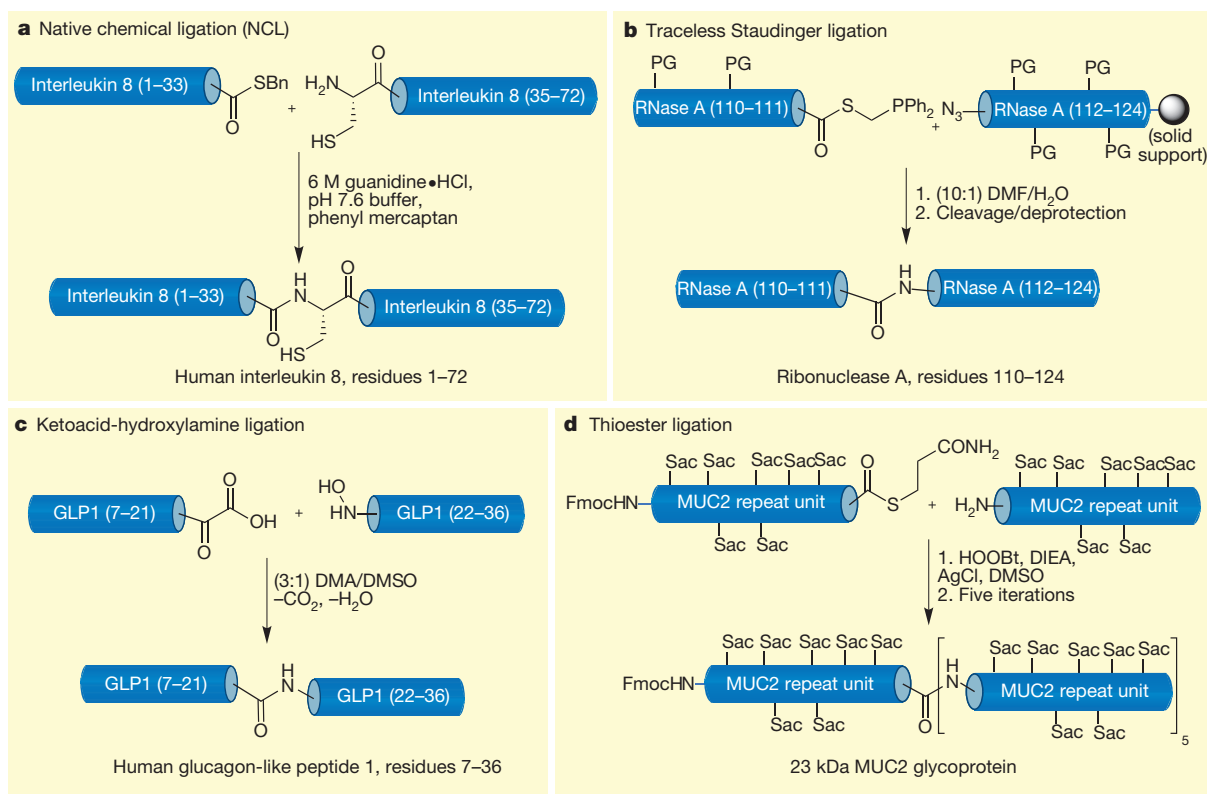


Figure 6 | Methods for chemoselective amide forming ligation for peptides, proteins and glycopeptides. **a**, Synthesis of human interleukin 8 by native chemical ligation (NCL) of a C-terminal peptide thioester and an N-terminal cysteine residue to give a native amide linkage. **b**, Traceless Staudinger ligation between a C-terminal phosphinothioester and an azide for the synthesis of a fragment of RNase A. **c**, Decarboxylative amide ligation between a C-terminal peptide α-ketoacid and an N-terminal hydroxylamine to form a native amide bond. In **a–c**, the numbers in parentheses represent the amino acid residues in

the N to C direction. **d**, Chemoselective silver-promoted 'thioester ligation' of a partially protected C-terminal peptide thioester and an N-terminal amine to obtain a fully synthetic 23 kDa MUC2 repeat glycoprotein. PG, protecting group; DMF, *N,N*-dimethylformamide; Fmoc, 9-fluorenylmethyloxycarbonyl; Sac, monosaccharide; HOOBt, 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine; DIEA, *N,N*-diisopropylethylamine; DMSO, dimethyl sulphoxide; MUC2, mucin 2 protein.

many proteins, ligation at serine or threonine residues has also gained attention. In a recent example, C-terminal *O*-salicylaldehyde ester was found to undergo ligation to give an amide with an N-terminal serine or threonine peptide with excellent reaction rates and selectivity⁵⁰.

The two most promising reactions to emerge from the efforts to identify new chemoselective ligations are the Staudinger⁵¹ ligation (Fig. 6b) and the α-ketoacid-hydroxylamine decarboxylative ligation (Fig. 6c)⁵². The traceless Staudinger amide-forming ligation, first reported in refs 53 and 54, involves the reaction of a peptidyl C-terminal phosphinothioester with a N-terminal peptide azide. This liberates a molecule of nitrogen and leads to an iminophosphorane intermediate, which breaks down to give the new amide bond. Several applications of this reaction have appeared, including the synthesis of RNase A^{55,56} in conjunction with NCL and the chemoselective formation of peptide macrocycles⁵⁷. Restrictions of the Staudinger ligation include its limited functional group tolerance, which necessitates partial or complete protection of peptide side chains.

The α-ketoacid-hydroxylamine ligation takes advantage of a selective reaction between a C-terminal peptide ketoacid and an N-terminal peptide hydroxylamine⁵⁸. The initial reaction between these two uniquely reactive functionalities is believed to produce a metastable hemiaminal intermediate that undergoes subsequent reaction to release a molecule each of carbon dioxide and water to produce a new amide bond. In a recent utilization of this ligation method, human glucagon-like peptide 1 (GLP1; also known as GCG), a 30-amino-acid therapeutic peptide, was synthesized in good yields and purity with fully unprotected peptide side chains⁵⁹. The highly chemoselective nature of the α-ketoacid-hydroxylamine ligation suggests an unconventional reaction mechanism,

which when fully identified could help in fine-tuning the reactive functionalities for developing new variants of this unusual ligation. An interesting off-shoot of the α-ketoacid-hydroxylamine ligation is the selective reaction between a α-ketophosphonic acid and *N*-(benzoyloxy)amine to generate an amide group in aqueous conditions without coupling reagents⁶⁰. Application of α-ketoacid-hydroxylamine ligation and its variants for the synthesis of α-peptides, β-peptides⁶¹ and glycopeptides⁶² have begun to appear.

Partially protected ligation strategies

Chemoselective amide-formation with partially protected fragments are especially useful for the assembly of complex homogeneous glycoproteins. In general, the syntheses of long glycopeptides with highly glycosylated side chains are difficult to prepare by conventional solid-phase chemistries when compared to peptides of similar length. To circumvent this problem, chemoselective glycoprotein ligation using relatively small and readily accessible glycopeptide fragments has emerged as the preferred strategy. In addition to many established and emerging methods⁶³, the silver-ion-mediated condensation of partially protected peptide amines with peptide thioacids or thioester, commonly known as the 'thioester ligation', has proven to be one of the most dependable methods for the synthesis of very large glycoproteins. Following preliminary reports on reactions of thioacids in ref. 35, it was shown⁶⁴ that thioesters could also be used for the ligation in the presence of an activating agent such as 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (HOOBt). In these methods, silver(I) reacts with the sulphur atom of a peptide thioacid or thioester to generate a reactive intermediate, which is then displaced by a nucleophilic amine leading

to a new amide bond. As a testimony to the power of thioester ligation, a 23 kDa basal MUC2 (mucin 2) tandem repeat glycoprotein containing 141 amino acids and 42 *N*-acetylgalactosamine units was synthesized⁶⁵. This represents one of the largest glycoproteins obtained by chemical synthesis (Fig. 6d). Although powerful for accessing large and complex glycoproteins, one of the drawbacks of the thioester ligation is the requirement of partial or complete protection of reactive functional groups in peptides or glycopeptides.

New reactions for amide-based biopolymers

Amide-based polymers have far-reaching applications, not only as materials encountered in everyday life but also as sophisticated tools for drug delivery, wound healing, biomineralization, and adhesives^{66,67}. Commercially produced polyamides are typically devoid of unprotected functionalities or complex stereochemistry that give rise to the unique and powerful properties of emerging classes of polyamides. Incorporating these elements, particularly in a controlled fashion, remains a considerable synthetic challenge.

Perhaps the most impressive progress to date has been the development of controlled/living polymerization methods⁶⁸ for the synthesis of polypeptides. In general, simple polypeptides can be accessed from α -amino acid-*N*-carboxyanhydrides, but sometimes they suffer from complex and uncontrollable side-reactions, leading to polypeptides with low molecular weight and high polydispersity index. These poor properties preclude the desirable higher order organizations frequently found in biomolecules, such as the quaternary structure of proteins. A seminal report disclosed⁶⁹ the first highly efficient organonickel initiator for the synthesis of co-polypeptides from amino-acid-derived α -amino acid-*N*-carboxyanhydrides that delivered low polydispersity, high-molecular-weight polymers (Fig. 7a). Several other organometal initiators for controlled/living polymerization are emerging, and the strategy has proven to be very reliable for the synthesis of high-molecular-weight polypeptides and polyglycopeptides⁷⁰. Recently, a completely different approach to the synthesis of polypeptides from non-amino-acid monomers was reported⁷¹. In this atom-economical method, imines and carbon monoxide undergo alternating copolymerization reactions in the presence of a cobalt catalyst to readily produce high-molecular-weight polypeptides (Fig. 7b). The advantage of this approach is the ready availability and low cost of the

starting materials, making it suitable for large-scale production of polypeptides.

Applications of new amide-forming reactions

What advances might these emerging methods for amide formation bring? Given the ubiquity of amides in both natural and unnatural materials, we cannot address all of the potential applications and instead focus on contemporary challenges or needs that may be addressed as these technologies mature.

Catalytic amide formations. Perhaps one of the greatest needs in amide-forming reactions are straightforward, catalytic approaches that reduce or eliminate the need for stoichiometric reagents. In addition to adding expense and chemical waste to a process, the toxicity and by-products associated with these reagents add complexity to otherwise routine transformations often employed in the synthesis of pharmaceuticals and agrochemicals. Boronic acids and related catalysts are likely to find use in specialized cases, but the high temperatures and unfavourable thermodynamics will probably limit their applicability. The redox-based methods employing either *N*-heterocyclic carbenes or metal catalysts proceed at lower temperatures and some do not require any added reagents. Some of these methods are likely to be adapted for the synthesis of non-peptidic amide-containing compounds.

Production of therapeutic peptides. The remarkable story of enfurvitide (Fuzeon), an anti-HIV agent, encapsulates both the success and inherent limitations of modern peptide synthesis⁷². The ambitious chemical production of Fuzeon by chemists at Roche using largely Fmoc solid-phase peptide synthesis serves as a landmark achievement in synthetic chemistry. But Fuzeon is a special case in that two, potentially problematic, fragment couplings of side-chain protected peptides occurred smoothly and the purification of the final peptide following global deprotection was relatively straightforward.

In practice, large-scale peptide manufacturing presents several challenges that are not well served by traditional solid-phase peptide synthesis⁷³. First and perhaps foremost is the difficulty of preparing highly pure materials devoid of impurities arising from additions, deletions or epimerization that occur during traditional peptide coupling. Late-stage fragment coupling requires the use of side-chain protected fragments, which can sometimes lead to poor yields, and still requires a global

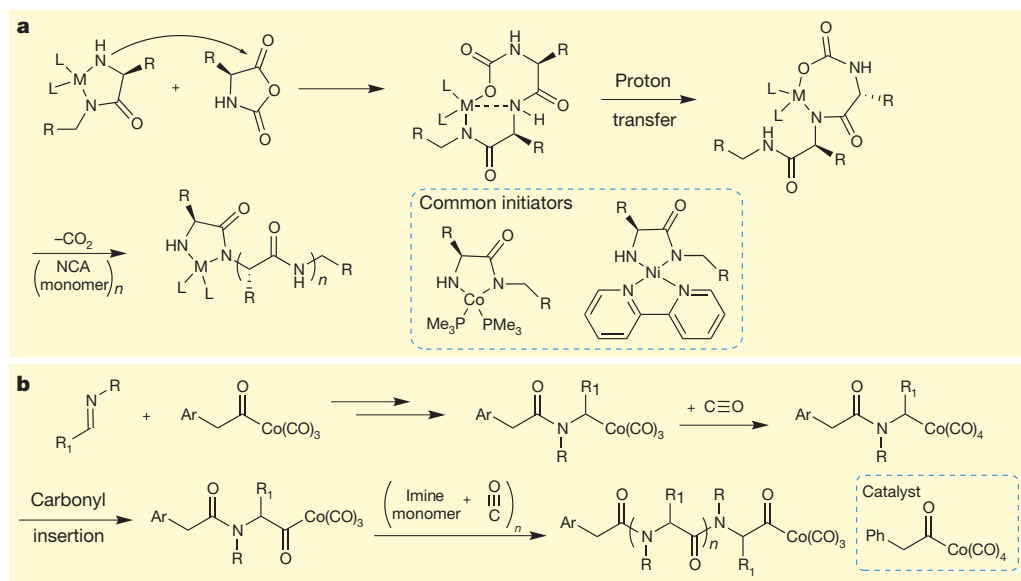


Figure 7 | Established and emerging methods for transition-metal-catalysed polyamide synthesis. **a**, Controlled/living polymerization of α -amino acid derived *N*-carboxyanhydride (NCA) with organometal initiator. Polymerization is initiated by reaction between NCA monomer (second structure from left) and the organo-metal initiator (leftmost structure) to generate an intermediate that undergoes a proton transfer and decarboxylation,

resulting in chain elongation and regeneration of the organo-metal initiator. **b**, Polypeptide synthesis from substituted imines and carbon monoxide, using a simple cobalt catalyst. Reaction between the imine and the cobalt-catalyst produces an amide intermediate, which undergoes carbonylation followed by carbonyl insertion to generate a reactive terminus for subsequent polymerization reactions. M, metal; L, ligand.

deprotection step. For these reasons, a fragment-coupling strategy is not always applicable for therapeutic peptide synthesis. A second challenge concerns the expense and chemical waste generated by iterative peptide couplings, which often require large excesses of protected amino acid monomers, coupling agents and additives. As an example, the synthesis of 1,000 kg of Fuzeon (containing 36 amino acids) requires about 45,000 kg of raw materials, not including the solvents used in the synthesis or purification⁷⁴.

In this regard, the evolving reactions for chemoselective amideligation offer great potential to improve the synthesis of therapeutic peptides in high purity and on a commercial scale. Ligations of fully unprotected peptide fragments offer considerable advantages over the coupling of fully or partially protected fragments, including obviating the need for post-coupling deprotections, the ability to work under aqueous conditions, and facile purification of the ligated product from any unreacted fragments. Ligation reactions are often suited for couplings at near equimolar ratios of reactant, whereas traditional, coupling-reagent-based approaches often require an excess of the carboxylic acid fragment. NCL comes closest to these ideals, but the requirement for a N-terminal cysteine or a cysteine surrogate invites the identification of new variants of this reaction and the continued development of novel ligation chemistries described above.

Synthesis of modified peptides. An important—and synthetically challenging—area is the identification and incorporation of modifications to α -peptide-based structures that increase metabolic stability, potency, bioavailability and permeability. Although most small or medium-sized peptides composed entirely of natural amino acids have limited stability, relatively modest modifications (including cyclization or the incorporation of backbone methyl groups) can dramatically improve their therapeutic index. For example, there is currently great excitement surrounding cyclotides, which are short peptides cyclized head to tail by an amide linkage and further cross-linked by disulphides (Fig. 8a)^{75,76}. The synthesis of such architectures, and cyclic peptides in general, by traditional methods is very challenging, as it often involves

fully protected peptides. This could prevent the formation of its native conformation for correct head to tail cyclization or disulphide formation. Advances in NCL have provided ways to synthesize diverse cyclic peptides and cyclotides. Conversely, new chemoselective ligation methods distinct from NCL could provide access to circular peptides and cyclotide-type peptide architectures that do not have a disulphide or a cysteine.

The most established class of modified peptides currently used as therapeutic agents are natural and unnatural *N*-methylated peptides, exemplified by cyclosporine A. *N*-methylation of single or multiple backbone amides has evolved as one of the favourite strategies to enhance potency, stability and structural rigidity. Synthetically, the increased steric demands of *N*-alkyl amino acids often cause great difficulties for amide bond formation (Fig. 8d)⁷⁷. They often require expensive coupling reagents and longer reaction times, risking epimerization and premature cleavage of the peptide from the resin by diketopiperazine formation. In this regard, new chemical processes that deliver *N*-methyl peptides, such as the reaction of thioacids with isonitriles followed by reduction, promise improved synthetic access to this class of compounds.

The urgency in devising new synthetic reactions for amide formation is enhanced by the isolation and structural identification of ever more complicated natural peptides with highly unusual modifications. Despite their unique biological activities, impressive chemical and metabolic stability, and complex architectures, none of these structures can currently be chemically synthesized, limiting opportunities for exploring the structure–activity relationship or developing designed amide structures with these enviable properties. Of particular interest is microcin J25, in which part of the peptide is threaded through a small macrocycle formed by an amide bond between a side chain and the N terminus (Fig. 8b)⁷⁸. These form compact, globular structures, ideally suited for mimicking protein–protein interactions^{79,80}. Subtilosin A⁸¹ and thuricin CD⁸² form another distinct class of naturally produced cyclic peptides, with multiple rings stitched together by a quaternary amino acid in which a cysteine sulphur is directly linked to the α -carbon of the peptide backbone (Fig. 8c). As traditional amide-forming reactions have largely failed for the synthesis of such structures⁸³, new amide-forming reactions must now move to the front line in these synthetic efforts.

Covalently modified synthetic proteins. Recent progress in the chemical synthesis of ‘oligomeric proteins’—consisting of several covalently linked protein subunits with biological relevance—illustrates the power of new amide-forming reactions to provide access to materials not readily available by either synthetic or biotechnological means. Of particular interest is ubiquitination, in which proteins are tagged for destruction by covalent post-translational modification of a lysine side chain by the attachment of multiple ubiquitin proteins. The application of a variant of NCL makes possible synthetic access to ubiquitinated proteins for investigations of fundamental biological questions^{84–86}.

A vital need in this area is a range of methods that would allow the chemical synthesis of much larger peptide structures, as current work is limited to perhaps at most 200 residues. Proteins in this size range can be obtained by NCL of two large fragments or sequences with orthogonal protecting groups or kinetically controlled ligations^{87,88}. One way of pushing this boundary further would be the identification of two distinct chemically orthogonal ligation methods that can operate together in the same flask, thereby allowing more than two fragments to be joined together in a controlled fashion. Such a strategy would reduce the number of purification, deprotection or manipulation steps necessary for the assembly of larger structures.

Functionalized amide-based polymers. Simple polyamides, such as nylon and related structures, have transformed the modern world by providing tough, durable and inexpensive materials for widespread applications. Despite these successes, chemists are far from matching naturally occurring highly versatile polyamides that have both structural as well as functional properties. For example, no chemical methods for the production of materials with the strength and properties of silks or collagens on a meaningful scale are currently available. The potential applications of

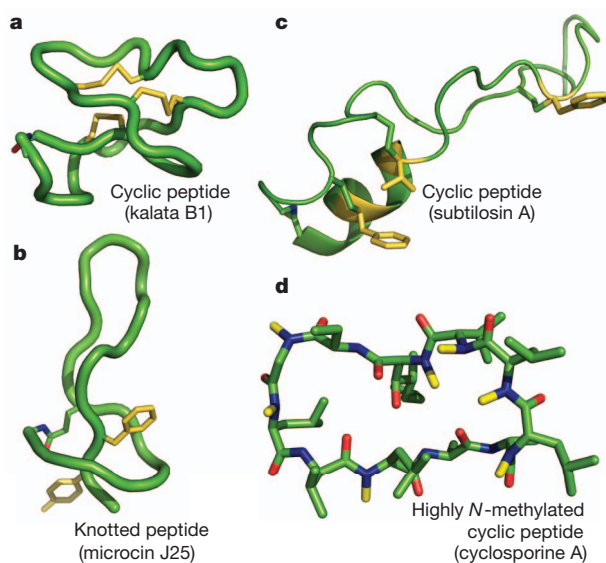


Figure 8 | Three-dimensional structures of representative examples for cyclic, knotted and *N*-methylated peptides. **a**, Kalata B1 is a cyclic peptide of the cyclotide family with six different loops formed by three disulphide bonds and a head to tail amide cyclization. **b**, Microcin J25 is a knotted peptide with unusually high stability imparted by threading the C terminus through a small macrocycle formed by the N terminus and a side-chain carboxylate. **c**, Subtilosin A represents a distinct class of bacterially derived cyclic peptide with three unusual side-chain sulphur to backbone α -carbon linkages and a head to tail cyclization. **d**, Cyclosporine A, a clinically important cyclic peptide with a very high proportion of *N*-methylated amides.

chemical methods for the controlled synthesis of richly functionalized co-block polyamides from simple starting materials are vast. Examples of the application of new amide-bond forming reactions to polyamide synthesis are already emerging. Milstein's ruthenium catalyst has been used⁸⁹ for direct synthesis of polyamides up to 30 kDa starting from simple diols and diamines, a process that has promise for low cost production of functionalized polyamides for biomaterials applications. In another impressive example, peptide self-assembly has been combined with chemical ligation to form⁹⁰ polypeptide fibrils and other architectures with molecular weights $\geq 3,000$ kDa. This is especially useful, as it provides access to synthetic polypeptides with a defined peptide sequence larger than any natural protein for applications in biophysics, basic protein science and cell biology.

Outlook

Over the past 100 years, synthetic chemists have advanced from the preparation of a dipeptide to the almost routine preparation of synthetic peptides of 50–100 or more amino acids. Structures of 20 to 40 amino acid residues can be prepared on a multi-kilogram scale. With a new generation of reagents and reactions, synthetic chemists will push further towards once unimaginable syntheses.

The adoption of catalytic methods for amide formation will have the most immediate impact on the contemporary preparation of amide-containing compounds. Such reactions will be both environmentally sustainable and cost effective. The first applications of these new reactions will be in the synthesis of small molecules, but the principles and underlying mechanisms can be extended to peptidic structures. This is particularly true of peptides containing unnatural amino acid residues, in which enantioselective variants of the catalytic reactions may provide direct access to not only the chiral starting material but the key amide linkage as well. The advantages of efficiency and selectivity inherent to catalytic methods for amide formation will offer new opportunities for enantioselective and chemoselective reactions, including amidations in the presence of reactive functional groups.

Improved amide-forming reactions that tolerate unprotected functional groups will continue the revolution in the preparation of synthetic proteins that began with NCL. Such reactions are needed to improve the production and discovery of the rapidly expanding class of peptide therapeutic agents currently under development. Such techniques will also aid the facile preparation of glycoproteins and other post-translationally modified structures. Among the long-term targets are methods for synthesis, deconstruction and reassembly of naturally expressed proteins. This would allow a portion of a protein of interest to be replaced by a synthetic sequence containing isotopically labelled residues for structural analysis, incorporation of unnatural amino acids for stability or enhanced activity, or specific probes for interrogating binding and biological function.

As we learn better ways to synthesize complex, highly functionalized amide-based structures without the need for aggressive reactants and cumbersome protecting groups, synthetic amides will provide a new generation of functional materials with controllable higher-order properties, such as adhesion, cell growth regulation and biomineralization. Until now, synthetic methods have not kept up with the promise of such materials. Chemists should take note of the challenges and opportunities existing in amide formation, and devise synthetic methods that are chemoselective, catalytic and sustainable.

- Arthur, G. *The Amide Linkage: Selected Structural Aspects in Chemistry, Biochemistry, and Materials Science* (Wiley-Interscience, 2000).
- Wieland, T. & Bodanszky, M. *The World of Peptides: A Brief History of Peptide Chemistry* (Springer, 1991).
- Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242 (2009).
- Coin, I., Beyermann, M. & Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nature Protocols* **2**, 3247–3256 (2007).
- Constable, D. J. C. *et al.* Key green chemistry research areas — a perspective from pharmaceutical manufacturers. *Green Chem.* **9**, 411–420 (2007).
- Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
- Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
Seminal paper describing the concept and utility of solid-phase peptide synthesis.
- Dawson, P. E., Muir, T. W., Clarklewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
Seminal paper describing the concept of native chemical ligation for protein synthesis.
- Pentelute, B. L., Gates, Z. P., Dashnau, J. L., Vanderkooi, J. M. & Kent, S. B. H. Mirror image forms of snow flea antifreeze protein prepared by total chemical synthesis have identical antifreeze activities. *J. Am. Chem. Soc.* **130**, 9702–9707 (2008).
- Mandal, K. *et al.* Racemic crystallography of synthetic protein enantiomers used to determine the X-ray structure of plectasin by direct methods. *Protein Sci.* **18**, 1146–1154 (2009).
- Torbееv, V. Y. & Kent, S. B. H. Convergent chemical synthesis and crystal structure of a 203 amino acid “covalent dimer” HIV-1 protease enzyme molecule. *Angew. Chem. Int. Edn* **46**, 1667–1670 (2007).
The largest fully chemically synthesized protein using native chemical ligation.
- Marchildon, K. Polyamides — still strong after seventy years. *Macromol. React. Eng.* **5**, 22–54 (2011).
- Marcelli, T. Mechanistic insights into direct amide bond formation catalyzed by boronic acids: halogens as Lewis bases. *Angew. Chem. Int. Edn* **49**, 6840–6843 (2010).
- Ishihara, K. Dehydrative condensation catalyses. *Tetrahedron* **65**, 1085–1109 (2009).
- Ishihara, K., Ohara, S. & Yamamoto, H. 3,4,5-Trifluorobenzeneboronic acid as an extremely active amidation catalyst. *J. Org. Chem.* **61**, 4196–4197 (1996).
The first highly active boronic acid catalyst for amide synthesis.
- Tang, P. Boric acid catalyzed amide formation from carboxylic acids and amines: n-benzyl-4-phenylbutyramide. *Org. Synth.* **81**, 262–268 (2005).
- Al-Zoubi, R. M., Marion, O. & Hall, D. G. Direct and waste-free amidations and cycloadditions by organocatalytic activation of carboxylic acids at room temperature. *Angew. Chem. Int. Edn* **47**, 2876–2879 (2008).
- Charville, H., Jackson, D., Hodges, G. & Whiting, A. The thermal and boron-catalysed direct amide formation reactions: mechanistically understudied yet important processes. *Chem. Commun.* **46**, 1813–1823 (2010).
- Bode, J. W. & Sohn, S. S. N-Heterocyclic carbene-catalyzed redox amidations of α -functionalized aldehydes with amines. *J. Am. Chem. Soc.* **129**, 13798–13799 (2007).
Details catalytic, redox neutral amide-forming reactions.
- Chiang, P.-C., Kim, Y. & Bode, J. W. Catalytic amide formation with α -hydroxyketones as acylating reagents. *Chem. Commun.* 4566–4568 (2009).
- Vora, H. U. & Rovis, T. Nucleophilic carbene and HOAt relay catalysis in an amide bond coupling: an orthogonal peptide bond forming reaction. *J. Am. Chem. Soc.* **129**, 13796–13797 (2007).
Details an organocatalytic oxidative method for amide formation.
- De Sarkar, S. & Studer, A. Oxidative amidation and azidation of aldehydes by NHC catalysis. *Org. Lett.* **12**, 1992–1995 (2010).
- Allen, C. L. & Williams, J. M. J. Metal-catalysed approaches to amide bond formation. *Chem. Soc. Rev.* **40**, 3405–3415 (2011).
- Ekoue-Kovi, K. & Wolf, C. One-pot oxidative esterification and amidation of aldehydes. *Chem. Eur. J.* **14**, 6302–6315 (2008).
- Tamaru, Y., Yamada, Y. & Yoshida, Z. Direct oxidative transformation of aldehydes to amides by palladium catalysis. *Synthesis* 474–476 (1983).
- Yoo, W.-J. & Li, C.-J. Highly efficient oxidative amidation of aldehydes with amine hydrochloride salts. *J. Am. Chem. Soc.* **128**, 13064–13065 (2006).
- Gunanathan, C., Ben-David, Y. & Milstein, D. Direct synthesis of amides from alcohols and amines with liberation of H₂. *Science* **317**, 790–792 (2007).
First report of the use of ruthenium catalysts for direct coupling of alcohols and amines for amide synthesis.
- Nordström, L. U., Vogt, H. & Madsen, R. Amide synthesis from alcohols and amines by the extrusion of dihydrogen. *J. Am. Chem. Soc.* **130**, 17672–17673 (2008).
- Muthaiah, S. *et al.* Direct amide synthesis from either alcohols or aldehydes with amines: activity of Ru(II) hydride and Ru(O) complexes. *J. Org. Chem.* **75**, 3002–3006 (2010).
- Chan, W.-K., Ho, C.-M., Wong, M.-K. & Che, C.-M. Oxidative amide synthesis and N-terminal α -amino group ligation of peptides in aqueous medium. *J. Am. Chem. Soc.* **128**, 14796–14797 (2006).
- Shen, B., Makley, D. M. & Johnston, J. N. Umpolung reactivity in amide and peptide synthesis. *Nature* **465**, 1027–1032 (2010).
- Li, X. & Danishefsky, S. J. New chemistry with old functional groups: on the reaction of isonitriles with carboxylic acids — A route to various amide types. *J. Am. Chem. Soc.* **130**, 5446–5448 (2008).
- Wu, X. Y., Stockdill, J. L., Wang, P. & Danishefsky, S. J. Total synthesis of cyclosporine: access to N-methylated peptides via isonitrile coupling reactions. *J. Am. Chem. Soc.* **132**, 4098–4100 (2010).
- Chatterjee, J., Gilon, C., Hoffman, A. & Kessler, H. N-Methylation of peptides: a new perspective in medicinal chemistry. *Acc. Chem. Res.* **41**, 1331–1342 (2008).
- Blake, J. Peptide segment coupling in aqueous-medium — silver ion activation of the thiolcarboxyl group. *Int. J. Pept. Protein Res.* **17**, 273–274 (1981).
- Rosen, T., Lico, I. M. & Chu, D. T. W. A convenient and highly chemoselective method for the reductive acetylation of azides. *J. Org. Chem.* **53**, 1580–1582 (1988).

37. Messeri, T., Sternbach, D. D. & Tomkinson, N. C. O. A novel deprotection/functionalisation sequence using 2,4-dinitrobenzenesulfonamide: Part 1. *Tetrahedr. Lett.* **39**, 1669–1672 (1998).
38. Messeri, T., Sternbach, D. D. & Tomkinson, N. C. O. A novel deprotection/functionalisation sequence using 2,4-dinitrobenzenesulfonamide: Part 2. *Tetrahedr. Lett.* **39**, 1673–1676 (1998).
39. Shangguan, N., Katukojvala, S., Greenberg, R. & Williams, L. J. The reaction of thio acids with azides: a new mechanism and new synthetic applications. *J. Am. Chem. Soc.* **125**, 7754–7755 (2003).
40. Yuan, Y., Chen, J., Wan, Q. A., Wilson, R. M. & Danishefsky, S. J. Toward fully synthetic, homogeneous glycoproteins: advances in chemical ligation. *Biopolymers* **94**, 373–384 (2010).
41. Wang, P. & Danishefsky, S. J. Promising general solution to the problem of ligating peptides and glycopeptides. *J. Am. Chem. Soc.* **132**, 17045–17051 (2010).
42. Rao, Y., Li, X. C. & Danishefsky, S. J. Thio FCMA intermediates as strong acyl donors: a general solution to the formation of complex amide bonds. *J. Am. Chem. Soc.* **131**, 12924–12926 (2009).
43. Talan, R. S., Sanki, A. K. & Sucheck, S. J. Facile synthesis of *N*-glycosyl amides using a *N*-glycosyl-2,4-dinitrobenzenesulfonamide and thioacids. *Carbohydr. Res.* **344**, 2048–2050 (2009).
44. Crich, D., Sana, K. & Guo, S. Amino acid and peptide synthesis and functionalization by the reaction of thioacids with 2,4-dinitrobenzenesulfonamides. *Org. Lett.* **9**, 4423–4426 (2007).
45. Crich, D. & Sasaki, K. Reaction of thioacids with isocyanates and isothiocyanates: a convenient amide ligation process. *Org. Lett.* **11**, 3514–3517 (2009).
46. Stephenson, N. A., Zhu, J., Gellman, S. H. & Stahl, S. S. Catalytic transamidation reactions compatible with tertiary amide metathesis under ambient conditions. *J. Am. Chem. Soc.* **131**, 10003–10008 (2009).
47. Rohde, H. & Seitz, O. Ligation–desulfurization: a powerful combination in the synthesis of peptides and glycopeptides. *Biopolymers* **94**, 551–559 (2010).
48. Blanco-Canosa, J. B. & Dawson, P. E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chem. Int. Edn* **47**, 6851–6855 (2008).
49. Fang, G.-M. *et al.* Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Edn* **50**, 7645–7649 (2011).
50. Li, X., Lam, H. Y., Zhang, Y. & Chan, C. K. Salicylaldehyde ester-induced chemoselective peptide ligations: enabling generation of natural peptidic linkages at the serine/threonine sites. *Org. Lett.* **12**, 1724–1727 (2010).
51. Staudinger, H. & Meyer, J. Über neue organische phosphorverbindungen III. Phosphinmethylester und phosphinimine. *Helv. Chim. Acta* **2**, 635–646 (1919).
52. Bode, J. W., Fox, R. M. & Baucum, K. D. Chemoselective amide ligations by decarboxylative condensations of *N*-alkylhydroxylamines and α -ketoacids. *Angew. Chem. Int. Edn* **45**, 1248–1252 (2006).
- Seminal paper describing the principle and application of the α -ketoacid-hydroxylamine ligation reaction.**
53. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A “traceless” Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2**, 2141–2143 (2000).
- Describes the development of the Staudinger reaction for peptide ligation.**
54. Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger ligation: a peptide from a thioester and azide. *Org. Lett.* **2**, 1939–1941 (2000).
- The development of traceless Staudinger ligation.**
55. Nilsson, B. L., Hondal, R. J., Soellner, M. B. & Raines, R. T. Protein assembly by orthogonal chemical ligation methods. *J. Am. Chem. Soc.* **125**, 5268–5269 (2003).
56. Köhn, M. & Breinbauer, R. The Staudinger ligation — a gift to chemical biology. *Angew. Chem. Int. Edn* **43**, 3106–3116 (2004).
57. Kleineweischede, R. & Hackenberger, C. P. R. Chemoselective peptide cyclization by traceless Staudinger ligation. *Angew. Chem. Int. Edn* **47**, 5984–5988 (2008).
58. Medina, S. I., Wu, J. & Bode, J. W. Nitron protecting groups for enantiopure *N*-hydroxyamino acids: synthesis of *N*-terminal peptide hydroxylamines for chemoselective ligations. *Org. Biomol. Chem.* **8**, 3405–3417 (2010).
59. Wu, J., Ruiz-Rodriguez, J., Comstock, J. M., Dong, J. Z. & Bode, J. W. Synthesis of human GLP-1 (7–36) by chemoselective α -ketoacid-hydroxylamine peptide ligation of unprotected fragments. *Chem. Sci.* **2**, 1976–1979 (2011).
60. Arora, J. S., Kaur, N. & Phanstiel, O. Chemoselective *N*-acylation via condensations of *N*-(benzoyloxy)amines and α -ketophosphonic acids under aqueous conditions. *J. Org. Chem.* **73**, 6182–6186 (2008).
61. Carrillo, N., Devalos, E. A., Russak, J. A. & Bode, J. W. Iterative, aqueous synthesis of β^3 -oligopeptides without coupling reagents. *J. Am. Chem. Soc.* **128**, 1452–1453 (2006).
62. Sanki, A. K., Talan, R. S. & Sucheck, S. J. Synthesis of small glycopeptides by decarboxylative condensation and insight into the reaction mechanism. *J. Org. Chem.* **74**, 1886–1896 (2009).
63. Hirano, K., Macmillan, D., Tezuka, K., Tsuji, T. & Kajihara, Y. Design and synthesis of a homogeneous erythropoietin analogue with two human complex-type sialyloligosaccharides: combined use of chemical and bacterial protein expression methods. *Angew. Chem. Int. Edn* **48**, 9557–9560 (2009).
64. Aimoto, S., Mizoguchi, N., Hojo, H. & Yoshimura, S. Development of a facile method for polypeptide-synthesis — synthesis of bovine pancreatic trypsin-inhibitor (BPTI). *Bull. Chem. Soc. Jpn* **62**, 524–531 (1989).
- The early application of the ‘thioester ligation’ method for protein synthesis.**
65. Hojo, H. *et al.* Chemical synthesis of 23 kDa glycoprotein by repetitive segment condensation: a synthesis of MUC2 basal motif carrying multiple O-GalNac moieties. *J. Am. Chem. Soc.* **127**, 13720–13725 (2005).
66. Deming, T. J. Synthetic polypeptides for biomedical applications. *Prog. Polym. Sci.* **32**, 858–875 (2007).
67. Rutledge, R. D. & Wright, D. W. *Biomimetic Synthesis of Peptides* (Wiley & Sons, 2009).
68. Ouchi, M., Terashima, T. & Sawamoto, M. Transition metal-catalyzed living radical polymerization: toward perfection in catalysis and precision polymer synthesis. *Chem. Rev.* **109**, 4963–5050 (2009).
69. Deming, T. J. Facile synthesis of block copolypeptides of defined architecture. *Nature* **390**, 386–389 (1997).
- The first organometal initiators for controlled polyamide synthesis.**
70. Kramer, J. R. & Deming, T. J. Glycopolypeptides via living polymerization of glycosylated-L-lysine *N*-carboxyanhydrides. *J. Am. Chem. Soc.* **132**, 15068–15071 (2010).
71. Sun, H., Zhang, J., Liu, Q., Yu, L. & Zhao, J. Metal-catalyzed copolymerization of imines and CO: a non-amino acid route to polypeptides. *Angew. Chem. Int. Edn* **46**, 6068–6072 (2007).
72. Bray, B. L. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nature Rev. Drug Discov.* **2**, 587–593 (2003).
73. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatsky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **15**, 40–56 (2010).
74. Marx, V. Roche’s fuzon challenge. *Chem. Eng. News* **83**, 16–17 (2005).
75. Ireland, D. C., Clark, R. J., Daly, N. L. & Craik, D. J. Isolation, sequencing, and structure’ activity relationships of cyclotides. *J. Nat. Prod.* **73**, 1610–1622 (2010).
76. Clark, R. J. *et al.* The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Edn* **49**, 6545–6548 (2010).
77. Rodriguez, H., Suarez, M. & Albericio, F. A convenient microwave-enhanced solid-phase synthesis of short chain *N*-methyl-rich peptides. *J. Pept. Sci.* **16**, 136–140 (2010).
78. Rebuffat, S. F. *et al.* Molecular knots as templates for protein engineering: the story of lasso peptides. *J. Pept. Sci.* **16**, 38 (2010).
79. Vincent, P. A. & Morero, R. D. The structure and biological aspects of peptide antibiotic microcin J25. *Curr. Med. Chem.* **16**, 538–549 (2009).
80. Knappe, T. A., Linne, U., Robbel, L. & Marahiel, M. A. Insights into the biosynthesis and stability of the lasso peptide capistrin. *Chem. Biol.* **16**, 1290–1298 (2009).
81. Kawulka, K. E. *et al.* Structure of subtilisin A, a cyclic antimicrobial peptide from *Bacillus subtilis* with unusual sulfur to α -carbon cross-links: formation and reduction of α -thio- α -amino acid derivatives. *Biochemistry* **43**, 3385–3395 (2004).
82. Rea, M. C. *et al.* Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc. Natl Acad. Sci. USA* **107**, 9352–9357 (2010).
83. Rosengren, K. J. *et al.* Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J. Am. Chem. Soc.* **125**, 12464–12474 (2003).
84. Kumar, K. S. A., Spasser, L., Erlich, L. A., Bavikar, S. N. & Brik, A. Total chemical synthesis of di-ubiquitin chains. *Angew. Chem. Int. Edn* **49**, 9126–9131 (2010).
85. Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nature Chem. Biol.* **6**, 750–757 (2010).
86. Yang, R., Pasunooti, K. K., Li, F., Liu, X.-W. & Liu, C.-F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem. Commun.* **46**, 7199–7201 (2010).
87. Tam, J. P., Yu, Q. & Yang, J.-L. Tandem ligation of unprotected peptides through thiopropyl and cysteinyl bonds in water. *J. Am. Chem. Soc.* **123**, 2487–2494 (2001).
88. Bang, D., Pentelute, B. L. & Kent, S. B. H. Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angew. Chem. Int. Edn* **45**, 3985–3988 (2006).
89. Zeng, H. & Guan, Z. Direct synthesis of polyamides via catalytic dehydrogenation of diols and diamines. *J. Am. Chem. Soc.* **133**, 1159–1161 (2011).
90. Ryadnov, M. G. & Woolfson, D. N. Self-assembled templates for polypeptide synthesis. *J. Am. Chem. Soc.* **129**, 14074–14081 (2007).

Acknowledgements J.W.B. is grateful for the support of diverse agencies for the development of new methods for amide formation, including the Arnold and Mabel Beckman Foundation, the David and Lucille Packard Foundation, Research Corporation for Science Advancement, the US National Institutes of Health (NIH-NIGMS) and the Swiss National Science Foundation (200021-131957). V.R.P. is supported by an ETH Fellowship. We thank all our past and present co-workers who have contributed to the discovery and development of new amide-forming reactions.

Author Contributions J.W.B. conceived the outline of this Review; J.W.B. and V.R.P. discussed, planned and wrote it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to J.W.B. (bode@org.chem.ethz.ch).

Cancer immunotherapy comes of age

Ira Mellman¹, George Coukos² & Glenn Dranoff³

Activating the immune system for therapeutic benefit in cancer has long been a goal in immunology and oncology. After decades of disappointment, the tide has finally changed due to the success of recent proof-of-concept clinical trials. Most notable has been the ability of the anti-CTLA4 antibody, ipilimumab, to achieve a significant increase in survival for patients with metastatic melanoma, for which conventional therapies have failed. In the context of advances in the understanding of how tolerance, immunity and immunosuppression regulate antitumour immune responses together with the advent of targeted therapies, these successes suggest that active immunotherapy represents a path to obtain a durable and long-lasting response in cancer patients.

The passive transfer of anticancer monoclonal antibodies and donor T cells in the context of allogeneic bone marrow transplantation are effective treatments for a variety of haematological and solid malignancies¹. Although not always thought of as ‘immunotherapy’, the success of these biotherapeutics probably reflects the ability of the donor cells or antibodies to induce an immediate immune reaction against the cancer, bypassing the requirement to activate endogenous immunity. These immune treatments have been well-established in oncology for several decades, and continued advances in antibody and T-cell engineering should further enhance their clinical impact in the years to come (Box 1).

In contrast to these passive immunotherapy strategies, the active stimulation of specific and durable antitumour immunity has proved elusive. In 1891, William Coley, a young New York surgeon, began intratumoral injections of live or inactivated *Streptococcus pyogenes* and *Serratia marcescens* in an effort to reproduce the spontaneous remissions of sarcomas observed in rare-cancer patients who had developed erysipelas². Given Elie Metchnikoff’s contemporaneous work demonstrating the immune system’s ability to cause inflammation and destroy invading bacteria, ‘Coley’s toxins’ made sense by acting to stimulate antibacterial phagocytes that might kill bystander tumour cells. Some significant responses were recorded over the ensuing 40 years, but successes were sporadic, difficult to reproduce and not obtained in a scientifically rigorous fashion. Superficial bladder cancer was one notable exception, for which the intravesical injection of live bacillus Calmette-Guérin after surgical resection was shown to prolong patient survival³. Other than this particular clinical setting, the approach was never embraced by oncologists who continued to rely on surgery and, increasingly, on effective new methods, such as radiation therapy and ultimately chemotherapy. Coley’s strategy was further discounted due to the very real risks associated with the administration of infectious, or at least pyrogenic, agents to already weakened cancer patients; this is ironic given the trauma associated with the treatments that did come into common use. Thus began the history of cancer immunotherapy. Before continuing, however, it is useful to summarize what must happen to elicit a protective immune response to cancer, and why overcoming these barriers has been so difficult.

Generating anticancer immunity is a multistep challenge

Based on our current understanding of the immune response, there are three distinct steps that must be achieved, either spontaneously or therapeutically, to mount effective antitumour immunity (Fig. 1). To

BOX 1

Established immune treatments

Nine monoclonal antibodies targeting six cancer-associated proteins (Her2/neu, EGFR, VEGF, CD20, CD52 and CD33) are approved for the treatment of solid and haematological malignancies. In addition to antagonizing oncogenic pathways, these biotherapeutics may act by opsonizing tumour cells and triggering their death or removal by antibody-dependent cellular cytotoxicity or phagocytosis⁹⁴. Ongoing investigations in murine models and patients raise the possibility that they may also stimulate adaptive immune responses in some settings⁹⁵. Recently, the successful conjugation of toxins to antibodies has been achieved, and these have induced a clinical response in patients who are refractory to the naked antibody⁹⁶. The concurrent administration of immunostimulatory cytokines such as IL-2 and GM-CSF may also enhance the efficacy of antibody therapy.

Allogeneic bone marrow transplantation and the infusion of donor lymphocytes can be a highly effective therapy for some leukaemias and lymphomas²⁴. The graft-versus-leukaemia effects involve the direct killing of tumour cells by donor lymphocytes, together with the subsequent induction of broader innate and adaptive reactions. On the basis of these clinical benefits, many groups are exploring the use of adoptive T-cell therapy in the autologous setting. Promising strategies include the use of lymphodepletion before T-cell infusion, and the engineering of new T-cell specificities with CARs⁹⁷.

Other immune treatments that have received the FDA approval include recombinant cytokines, such as IL-2 (Proleukin), which is used for melanoma and renal cell cancer. Response rates are low (~15%) and the significant risk of serious systemic inflammation requires administration as an in-patient. Interferon- α is another agent that gained approval for ‘immunological cancers’ (that is, melanoma or renal cell cancer). Although also associated with low response rates and high-dose toxicity, a small subset of melanoma patients, who are also predisposed to autoimmunity, has been shown to exhibit an impressive survival response⁹⁸. It has been, however, difficult to pre-identify these patients, which limits the use of the approach. Yet, when seen, responses are durable, suggesting they reflect active antitumour immunity.

¹Genentech, 1 DNA Way, South San Francisco, California 94080, USA. ²Ovarian Cancer Research Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. ³Department of Medical Oncology and Cancer Vaccine Center, Dana-Farber Cancer Institute, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

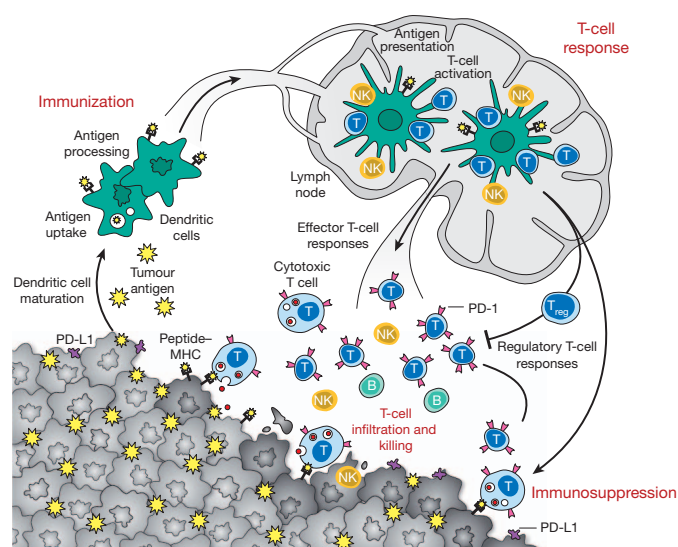


Figure 1 | Generation and regulation of antitumour immunity.

Understanding the events in generating and regulating antitumour immunity suggests at least three sites for therapeutic intervention: promoting the antigen presentation functions of dendritic cells, promoting the production of protective T-cell responses and overcoming immunosuppression in the tumour bed. Antitumour immune responses must begin with the capture of tumour-associated antigens by dendritic cells, either delivered exogenously or captured from dead or dying tumour cells. The dendritic cells process the captured antigen for presentation or cross-presentation on MHC class II and class I molecules, respectively, and migrate to draining lymph nodes. If capture and presentation occurred in the presence of an immunogenic maturation stimulus, dendritic cells will elicit anticancer effector T-cell responses in the lymph node; if no such stimulus was received, dendritic cells will instead induce tolerance leading to T-cell deletion, anergy or the production of T_{reg} cells. In the lymph node, antigen presentation to T cells will elicit a response depending on the type of dendritic cell maturation stimulus received and on the interaction of T-cell co-stimulatory molecules with their surface receptors on dendritic cells. Thus, interaction of CD28 or OX40 with CD80/86 or OX40L will promote potentially protective T-cell responses, while interaction of CTLA4 with CD80/86 or PD-1 with PD-L1/PD-L2 will suppress T-cell responses, and possibly promote T_{reg} formation. Antigen-educated T cells (along with B cells and NK cells) will exit the lymph node and enter the tumour bed, where a host of immunosuppressive defence mechanisms can be produced by tumours (or infiltrating myeloid cells) that oppose effector T-cell function. These include the upregulation of PD-L1/L2 on the cancer cell surface, release of PGE₂, arginase and IDO (all T-cell suppressors), and the release of VEGF (triggered in part by intratumoural hypoxia), which inhibits T-cell diapedesis from the vasculature, and thus infiltration into the tumour bed.

initiate immunity, dendritic cells must sample antigens derived from the tumour, which can be ingested *in situ* or delivered exogenously as part of a therapeutic vaccine. These antigens might reflect one or more of the many mutated proteins that are typical of cancer, the products of non-mutated genes that are preferentially expressed by cancer cells (for example, cancer-testis antigens), or differentiation antigens associated with the cancer's tissue of origin, but against which thymic or peripheral tolerance has not been completely established (for example, melanosome-associated proteins in melanoma)^{4,5}. On antigen encounter, the dendritic cells would also have to receive a suitable activation ('maturation') signal, allowing them to differentiate extensively to promote immunity as opposed to tolerance including enhanced processing and presentation of tumour-antigen-derived peptides^{6,7}. Activation signals could be therapeutically supplied exogenously (for example, Toll-like receptor (TLR) ligands or agonist antibodies against activating receptors such as CD40) or endogenously: dying or necrotic tumour cells release factors (for example, high mobility group proteins or ATP) that are thought to result in the immunogenic maturation of dendritic cells. In addition, tumour cells seem to express resident endoplasmic reticulum proteins

ectopically on the plasma membrane (for example, calreticulin) that promote their phagocytosis, possibly enabling the presentation of tumour antigens on major histocompatibility complex (MHC) class I and class II molecules. Some forms of chemotherapy or targeted therapy may promote a more immunogenic phenotype⁸.

Next, in lymphoid organs, tumour-antigen-loaded dendritic cells must generate protective T-cell responses⁹. The precise type of T-cell response needed is unknown, but they certainly include the production of CD8⁺ effector T cells with cytotoxic potential. Dendritic cells may also trigger antibody and natural killer (NK) or natural killer T (NKT)-cell responses, which may contribute to tumour immunity. The lymph node is a second potential site for therapeutic intervention, providing agents that may help guide the T-cell response. However, the dendritic cells are again key players because they must have been matured by a stimulatory adjuvant to have a chance at eliciting the desired T cells. Presentation of antigens by dendritic cells at the steady state (that is, dendritic cells that have not received an immunogenic maturation signal) promotes tolerance by regulatory T cell (T_{reg}) production^{10–13}, which would oppose an antitumour response.

BOX 2

Mechanisms of immune suppression

Tumours escape immune attack by a variety of complementary mechanisms of immunosuppression, many of which operate in parallel. Among paracrine mediators, adenosine, prostaglandin E₂, TGF- β and VEGF-A exert multiple direct and indirect immunosuppressive activities. These mediators may function in the suppression of dendritic cells, indirectly inhibiting T-cell penetration into the tumour bed or directly suppressing effector T-cell activation while enhancing the function of T_{reg} cells. For example, adenosine, which is released by tumour cells under hypoxia, contributes to the suppression of T-cell activation and T_{reg} expansion. VEGF-A can also suppress proper T-cell development and function: VEGF-A treatment of mouse splenocytes during T-cell stimulation was shown to induce IL-10 production from T cells while suppressing IFN- γ production⁹⁹. Tumour cells can also directly escape T-cell recognition by downregulating MHC class I or by disabling other components of the antigen processing machinery. Shedding of soluble NKG2D ligands such as MIC-A or MIC-B can severely compromise the ability of effector T cells to function in the tumour microenvironment. In addition, tumour cells may upregulate surface ligands, which mediate T-cell anergy (or exhaustion), including PD-L1 and other ligands to inhibitory T-cell receptors. A variety of leukocyte subsets infiltrating tumours are also able to suppress T-cell function. In addition to T_{reg} cells (the accumulation of which in tumours correlates with poor prognostic outcome¹⁴), other suppressive lymphocyte subsets have been reported including IL-10 producing B cells and B regulatory cells, type II NKT cells, NK cells and $\gamma\delta$ T cells. Myeloid lineage cells also promote immune suppression in tumours. Among these are mainly the poorly understood myeloid-derived suppressor cells (MDSC)¹⁰⁰. Their characterization is ultimately based on their ability to suppress T and NK cell activation, probably through several mechanisms including nitric oxide, reactive oxidative species, arginase, IL-10 and TGF- β ; there are also reports that MDSCs may specifically induce the expansion of T_{reg} cells. Tumour stroma cells have an important immune modulatory role. The so-called cancer-associated fibroblasts can promote the recruitment and function of immunosuppressive cells through the secretion of CCL2 and CXCL12. In addition, they can suppress effector T-cells through secretion of TGF- β . Finally, myeloid-derived mesenchymal stem cells exert important immunosuppressive functions by blocking proliferation and function of T effector cells. Further study is needed to determine which of these mechanisms are most important in general, and which determine the immune status in individual patients.

Finally, cancer-specific T cells must enter the tumour bed to perform their function at which point immune suppression becomes a challenge (Box 2). Tumours may (presumably by skewing dendritic cell maturation) prevent immunization, trigger the 'wrong' immune response or enable the local accumulation or expansion of T_{reg} cells that would oppose the activity of effector T cells. Indeed, infiltration of T_{reg} cells correlates with poor prognosis in a variety of epithelial tumour types^{14,15}. Tumours may downregulate their expression of MHC class I molecules or their expression of target tumour antigens and can also produce a variety of surface molecules (for example, PD-L1 or PD-L2) that engage receptors on the surfaces of activated T cells (PD-1), causing T-cell anergy or exhaustion^{16,17}. Expression of such suppressive ligands can be associated with oncogenic mutations seen in many cancers (for example, *PTEN* loss)¹⁸. Additionally, tumours can release immunosuppressive molecules, such as indoleamine 2,3-dioxygenase (IDO), which consumes tryptophan and limits T-cell function^{19,20}. Myeloid-derived suppressor cells can also be recruited into the tumour, and release additional T-cell suppressors, such as arginase and nitrous oxide synthase²¹. Hypoxia in the tumour microenvironment may promote the generation of adenosine, which inhibits effector T-cell function in a similar way²². Hypoxia can also lead to the production of CCL28, which attracts immigration of T_{reg} cells²³. Finally, tumour stroma cells can also suppress the function of effector lymphocytes. For example, mesenchymal stem cells block proliferation and function of effector T cells²⁴, whereas tumour vascular cells can suppress T-cell adhesion (to tumour endothelium) and prevent homing to tumours. This effect is, in part, mediated by vascular endothelial growth factor (VEGF)²⁵ as well as by the endothelin-B receptor (ETBR, also known as EDNRB)²⁶.

Thus, the success of immunotherapy seems unlikely. Any approach would have to overcome several significant barriers: that tumour-associated antigens are typically closely related or identical to self antigens, that it will be very difficult to separate therapeutic responses from pathological autoimmune responses, that both central and peripheral tolerance would have conspired to deplete or inactivate the relevant T-cell repertoire, and that the tumour microenvironment is inherently immunosuppressive. However, there now seems to be a path to clinical success.

Cancer vaccines are finally showing early signals of activity

Vaccines come in two formats: prophylactic and therapeutic⁹. Prophylactic (or preventative) vaccines have been used with considerable success for the prevention of cancers of viral origin, such as hepatitis B virus and human papillomavirus (HPV), where the aetiological agent is known. In contrast, the development of therapeutic vaccines to treat existing disease has proved problematic. The long history of failure has tainted the entire strategy of immunotherapy in the eyes of many oncologists.

The idea of a therapeutic cancer vaccine originated with the discovery that patients can harbour $CD8^+$ and $CD4^+$ T cells specific for cancer-testis or differentiation antigens expressed in their tumours⁴. Vaccination might reasonably be expected to amplify the frequency and strength of these pre-existing responses or perhaps induce some *de novo* reactions. Additionally, clinicopathological studies have demonstrated a strong association between prolonged patient survival and the presence of intratumoral $CD3^+$ or $CD8^+$ cytotoxic T cells and an interferon- γ (IFN- γ) gene signature^{27,28}. Thus, if vaccination could trigger these types of T-cell responses, then a clinical benefit might be expected.

Unfortunately, the many initial attempts were compromised by a poor understanding of the mechanism of immunization, specifically the role of dendritic cells. Mostly conducted by academic groups, thousands of patients were treated with vaccines consisting of short peptides, often without an effective dendritic-cell-activating adjuvant²⁹. Free peptides are likely to have poor pharmacokinetic properties and may be rapidly cleared before being loaded onto dendritic cells, where their half-life may also be short. Without an adjuvant, the dendritic cells might remain in the steady state and promote tolerance, as much as immunity. As a result, there was typically poor immunization, infrequent response to the selected tumour antigens (assuming they were even the correct ones) and

minimal therapeutic benefits. Recently, however, the co-administration of interleukin (IL)-2 as an immune stimulant with a short peptide derived from glycoprotein 100 (gp100), a melanocyte differentiation antigen, was shown, under some conditions, to augment tumour responses and prolong progression-free survival compared with IL-2 alone in advanced melanoma patients³⁰. These findings indicate that increasing immune activation with peptide vaccines is a critical step for improving therapeutic efficacy.

Because the importance and function of dendritic cells in stimulating T-cell responses is now well known, current vaccine trials are designed more rationally. One potentially promising approach involves the use of peptides (~20-mer) that are somewhat longer than those that bind to MHC class I molecules (10–12-mer). In the presence of a suitable dendritic-cell-activating adjuvant, these peptides are thought to be more efficient at generating effector T cells, perhaps because some processing may be required. A recent study of peptides derived from the HPV-16 E6 and E7 viral oncoproteins administered in incomplete Freund's adjuvant showed clinical responses in 15 of 19 women with vulvar intraepithelial neoplasia³¹. Tumour regressions were associated with the generation of HPV-specific, IFN- γ -producing $CD4^+$ and $CD8^+$ T cells. These favourable results might reflect in part the selection of viral gene products for immunization, because these proteins might be more readily recognized as foreign by the host. Indeed, a small trial of long peptides derived from TP53 (previously known as p53), a tumour suppressor often mutated in cancer, delivered in Montanide (an emulsion-adjuvant) induced a weaker IFN- γ -producing T-cell-response and no tumour regressions in advanced ovarian cancer patients, indicating the need to optimize these formulations further, or that p53 is simply not sufficiently immunogenic in this setting; in contrast to viral antigens, p53 might also have to break pre-existing tolerance to self³².

Full-length proteins are also being explored as targets for cancer vaccinations, as they contain a broader profile of epitopes that might be presented by dendritic cells. Among these approaches, GlaxoSmithKline is currently conducting a large (>2,500 patients) randomized phase III trial using a recombinant fusion protein encoding a single cancer-testis antigen (MAGE-A3) in HLA-A2-positive non-small cell lung cancer patients, together with their ASO2B adjuvant consisting of a saponin/lipid-A emulsion combined with TLR4 and TLR9 agonists. Initial read-outs from the phase II trial (180 patients) showed some survival response (27%), but this did not reach statistical significance³³. Objective T-cell responses were not reported, nor were data on the level or homogeneity of MAGE-A3 expression in the study group; this latter parameter might be especially important, as immune attack on sub-populations of lung cancer cells that lack MAGE-A3 would require diversification of the T-cell response to additional cancer antigens.

Another target that has attracted some attention is the antigen receptor on B-cell lymphomas (idiotype), which is an example of a clonally expressed tumour-specific antigen. Three randomized phase III trials testing recombinant idiotypes (prepared individually for each patient) administered with granulocyte-macrophage colony-stimulating factor (GM-CSF) to attract dendritic cells were undertaken. One trial, in which eligibility was restricted to subjects who first achieved a complete response to cytotoxic chemotherapy, indicated that vaccination might prolong progression-free survival. However, the two other studies failed to reveal a clinical benefit, which might reflect differences in patient populations, vaccine manufacturing methods or that the approach is not sufficiently robust³⁴.

Viral vectors encoding tumour antigens are another vaccine platform undergoing evaluation. These strategies exploit the strong immune response directed against viral components to enhance reactivity against the cancer antigen. One such phase II trial (125 patients), conducted by Bavarian-Nordic and the United States National Cancer Institute, involved an initial inoculation of a recombinant vaccinia virus encoding prostate-specific antigen and the adhesion molecules B7-1, ICAM-1 and LFA-3; a similarly configured fowlpox vector was administered subsequently in a prime-boost strategy, and GM-CSF was administered with the vectors for further immune stimulation (the entire vaccine

product is termed PROSTVAC). In addition to reactivity against the pox viral gene products, the introduction of adhesion molecules into the infected cells was intended to have them serve as surrogate dendritic cells, although the specializations of dendritic cells for T-cell stimulation extend well beyond these three adhesion molecules. Whereas vaccination had no impact on progression-free survival, there was an overall survival benefit: 25.1 months versus 16.6 months in the control group (patients treated with empty vector plus saline)³⁵. A larger phase III trial is planned using vector alone, vector plus GM-CSF (as a dendritic cell adjuvant or attractant) or empty vector plus GM-CSF.

A variety of earlier stage trials are also in progress, using defined antigens that are delivered to dendritic cells by coupling to dendritic-cell-targeted monoclonal antibodies (for example, DEC-205) or further peptide and viral vector trials, in conjunction with various adjuvants. The ability of these approaches to facilitate durable anticancer immune responses, however, remains to be demonstrated, particularly on their own.

Another strategy for vaccine therapy involves the use of cell-based approaches. One of the ideas underlying this strategy is that an actual cancer cell would already harbour a wide range of tumour-associated antigens (including mutant proteins), so that if used as a vaccine the problem of antigen selection would be reduced. A meta-analysis of 173 published peer-reviewed immunotherapy trials in various solid tumour types revealed that patients immunized with whole-tumour antigen had low, but significantly higher rates of objective clinical response (8.1%) than patients immunized with molecularly-defined tumour antigens (3.6%)³⁶. Although autologous tumour cells are the best choice of immunogen for this approach, the complexities of vaccine manufacture for individual patients has led to the application of allogeneic tumour cell lines. Among these strategies, the one that progressed furthest into clinical development was GVAX for prostate cancer, advanced by Cell Genesys (no longer exists). The vaccine product was comprised of two prostate cancer cell lines that were stably engineered to secrete GM-CSF. After early promising results, GVAX failed in phase III trials due to lack of clinical efficacy³⁷. This failure might reflect a lack of sufficient immunogenicity of the approach, alterations in preparation of the vaccine product necessitated by commercial scale-up or the inability of allogeneic tumour cell lines to represent adequately the spectrum of antigens characteristic of individual prostate cancers in patients.

There has also been considerable interest, particularly among academic groups, in developing dendritic-cell-based vaccines. In this approach, dendritic cells are isolated from a cancer patient, loaded with antigens (peptides or even tumour cell lysates) *ex vivo*, activated and then re-infused back into the patient³⁸. Although there have been some promising hints of efficacy and response, as yet no clearly positive studies have been reported and the approach has not gained broad support from the biotech-pharmaceutical industry, given the complexities of cell isolation, *ex vivo* manipulation and re-infusion. There is, however, one notable exception, which will be discussed shortly.

In general, there are many barriers to success, or at least to quantifying the success, of cancer vaccines administered as 'single agents'. First, the criteria for defining optimal tumour antigens remain to be fully defined^{39,40}. Mere expression in the target tumour population may be inadequate for predicting the ability to generate protective T-cell response. Identification of peptides bound to MHC class I on the tumour by mass spectroscopy can identify those antigens that yield potential targets⁴¹, but these peptide-MHCI complexes might not be sufficiently immunogenic rejection antigens (that is, able to generate effective T-cell responses). Moreover, antigen expression within a tumour bed can be heterogeneous. Even the allogeneic tumour cell line based approaches are limited by the fact that a given patient's tumour will probably harbour mutations that are not found in the vaccine product. Second, the optimal adjuvant for producing antitumour CD8⁺ T-cell responses that can be used safely and effectively in humans is not yet clear. The desired adjuvant (or adjuvant combination) will be one capable of triggering the maturation of dendritic cells to a state where they can facilitate the generation of tumour-reactive, CD8⁺ cytotoxic T cells. Finally, although it is likely that

conditions for immunization will eventually be optimized, the effectiveness of a tumour-specific T-cell population may still be limited by the multiple mechanisms of immunosuppression used by tumours to guard against T-cell killing. These are not reasons to eliminate a vaccine arm from consideration as part of immunotherapy, but rather to highlight some of the difficulties in assessing success in the absence of other immunomodulatory agents. In the same way as other forms of targeted therapy for cancer, the discovery and application of predictive biomarkers or diagnostics, which could identify those patients most likely to benefit from a given vaccine, will be an important challenge for future development.

An efficacious cell-based vaccine for prostate cancer?

A fact of drug development is that performance in the clinic is the final arbiter of success. The first validation of active immunotherapy as a viable approach to cancer treatment was the Food and Drug Administration (FDA) approval in April 2010 of Provenge (sipuleucel-T) for advanced prostate cancer. Provenge was originally assumed to be an autologous dendritic-cell-based vaccine⁴² (it was developed by the eponymously named company, Dendreon); however, it actually comprises an incompletely characterized, complex mixture of peripheral blood mononuclear cells supplemented with a cytokine and tumour-derived differentiation antigen. In the pivotal phase III trial⁴³, total peripheral blood mononuclear cells were collected from patients by leukopheresis at 0, 2 and 4 weeks and then cultured for 36–44 h at 37 °C in medium containing a fusion protein composed of prostatic acid phosphatase (PAP, a tumour-associated differentiation antigen) and GM-CSF before reinfusion. This was compared with a placebo that was prepared by incubating one-third of the leukopheresis product for 36–44 h at 2–8 °C without exposure to the fusion protein (the remaining two-thirds were cryopreserved to allow for subsequent vaccine manufacture in a salvage protocol). Thus, the control not only lacked the presumptive immunizing antigen, but was also processed differently.

The clinical results showed little evidence of tumour shrinkage or delay in disease progression. By standard response evaluation criteria in solid tumours (RECIST) criteria, only 1 out of the 341 patients in the active arm showed a partial response. At least a 50% reduction in prostate-specific antigen levels on at least two visits was shown in 2.6% of the patients versus 1.3% in the placebo group. Nevertheless, a 4.1 month improvement in median survival was achieved (25.8 months versus 21.7 months), which was deemed significant by the FDA in a patient population that has few, if any other, effective therapeutic options.

While Provenge is clearly a cell-based therapy, there may be other mechanisms involved. Although the majority (66%) of survivors showed an antibody response to the fusion protein, the fraction of patients producing antibodies that recognized endogenous PAP was much lower (28.5%). Moreover, T-cell responses to either the fusion protein or PAP were not associated with survival. These discrepancies might reflect a limitation of monitoring antitumour immune responses in the peripheral blood compared with the tumour microenvironment. However, they also raise the possibility that other undefined factors in the cellular product may have an important role. Further studies are required to understand the therapeutic mechanism of Provenge, and to define the impact of the different cell-processing procedures on the placebo product.

The lack of tumour shrinkage, the criterion typically used to gauge the efficacy of cancer treatments, in the face of a survival benefit is surprising, but perhaps not unexpected for immunotherapy. As seen pre-clinically, an effect on pre-existing tumours due to immune manipulations can be delayed while an immune response develops^{44,45}. Furthermore, biopsies of metastases after vaccination in some clinical trials revealed the presence of immune infiltrates that mediate tumour destruction in association with extensive oedema, which may be followed by fibrosis⁴⁶. These histopathological findings suggest that monitoring tumour size alone may be inadequate for assessing the overall therapeutic effects of vaccination. As discussed later, these considerations apply to the evaluation of CTLA-4 antibody blockade, highlighting the need to modify

tumour response criteria in light of new insights into the biology of immunotherapy⁴⁴.

The potential of directly harnessing lymphocytes

Adoptive transfer of lymphocytes with tumoricidal properties can, in theory, bypass the daunting task of breaking tolerance to tumour antigens and generating a high frequency of high avidity effector T cells. The discovery that host lymphodepletion facilitates engraftment of adoptively transferred T cells has enabled the successful transfer of *ex-vivo*-expanded tumour-infiltrating lymphocytes (TILs) from patients with melanoma, with marked clinical responses, some of which are complete and durable^{47,48}.

Advances in T-cell culturing methods and T-cell engineering, through retroviral vectors carrying cloned T-cell receptors or chimaeric antigen receptors (CARs) enriched by co-stimulatory signalling domains, has expanded the opportunities for adoptive T-cell therapy beyond patients with resectable tumours harbouring reactive T cells to a larger population with solid tumours expressing the cognate target(s)⁴⁹. The advent of CARs bypasses the need for tumour cells to possess functional antigen-processing machinery and express antigen through MHC class I or II molecules; transduced T cells can recognize the intact surface protein through the affinity domain (usually a scFv antibody) of the artificial CAR. Early clinical results appear very promising^{50,51}. However, safety issues surrounding the selection of the target, the paucity of such targets, manufacturing complexities and costs, and the lack of durable response in many patients indicates that additional interventions are required to properly direct and activate T cells in the tumour microenvironment.

A far more convenient approach may be the therapeutic use of bispecific antibodies that engage both the TCR and an antigen on the tumour cell surface. In B-cell lymphoma and leukaemia, MicroMet⁵² has applied this approach with impressive clinical success. However, the platform has a number of liabilities including neurotoxicity and the necessity for continuous pump-mediated infusion due to the very rapid clearance of the antibody fragment. Nevertheless, such bispecific antibodies further emphasize the possibility of repurposing T-cell specificity for therapeutic benefit.

Ipilimumab's emergence as an effective therapy

The most important development for cancer immunotherapy, and hopefully for the benefit of many cancer patients, was the recent readout of the ipilimumab phase III trials in late-stage metastatic melanoma. Not only was a clear survival advantage observed for a patient group with no other therapeutic options, but it was achieved with an agent whose mechanism of action is virtually certain to involve the modulation of endogenous T-cell responses. The results were deemed significant enough that in March 2011 the FDA granted broad approval for use in patients with metastatic melanoma, either as initial therapy or after relapse.

Ipilimumab is a monoclonal antibody to CTLA4, whose role in regulating T-cell function has been studied for many years by a number of groups, notably by Allison and colleagues⁵³. CTLA4 is a key negative regulator that is recruited to the plasma membrane on T-cell activation where it binds to members of the B7 family of accessory molecules expressed by dendritic cells and other antigen-presenting cells (Fig. 2). CTLA4 ligation effectively inhibits further activation and expansion, thereby controlling the progress of an immune response and attenuating the chances for chronic autoimmune inflammation. The negative regulation is overcome by use of a blocking antibody. The fundamental importance of CTLA4 in controlling T-cell function is well-illustrated by the phenotype of *Ctla4*^{-/-} mice, which die of an aggressive lymphoproliferative disorder at a young age^{54,55}. Interestingly, CTLA4 ligation is also important for the immune suppressive function of T_{reg} cells, by further assisting to dampen T-cell responses⁵⁶. T_{reg} function is also thought to be blocked by anti-CTLA4.

The rationale for using anti-CTLA4 in cancer therapy was to unrestrain pre-existing anticancer T-cell responses and possibly trigger

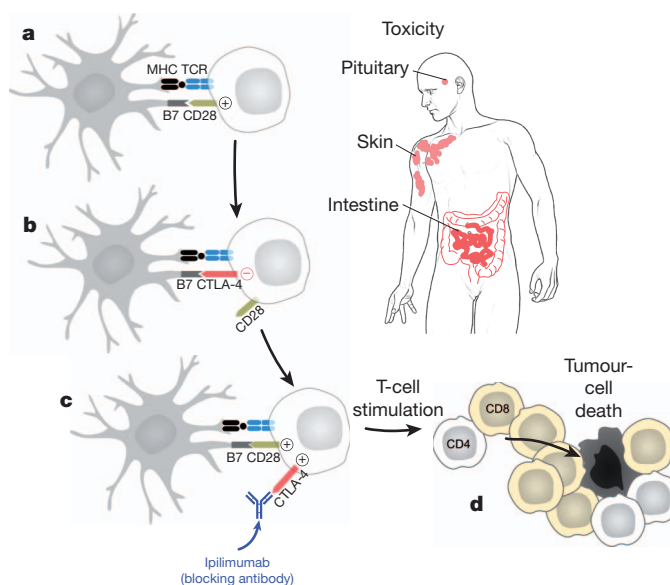


Figure 2 | Biological activities of CTLA-4 antibody blockade. **a**, On encountering a dendritic cell presenting a cognate tumour-antigen-derived peptide epitope and expressing B7 co-stimulatory molecules (CD80, CD86), specific antitumour T cells become activated through TCR and CD28 signalling. **b**, CTLA-4 is subsequently upregulated and preferentially engages B7 to attenuate T-cell response. **c**, Ipilimumab blocks CTLA-4 function, thereby allowing enhanced T-cell stimulation and a more potent antitumour reaction. Ipilimumab may also antagonize CTLA-4 on regulatory T cells to limit their ability to suppress the antitumour T-cell effector response (not shown). **d**, CTLA-4 antibody blockade compromises tolerance to some normal tissue antigens, provoking inflammatory toxicities that can have an impact on the skin, pituitary gland and intestine in human patients.

new responses. It is well-known for melanoma (and other diseases) that TILs exist, and they can bear specificity for tumour antigens^{4,57–59}. Pre-clinical studies using mouse models were promising, which led two companies (Pfizer and BMS/Medarex) to put two different anti-CTLA4 antibodies into the clinic. Phase II trials failed to reach their endpoints of tumour regression, but BMS/Medarex felt there was sufficient potential for long-term benefit, and so a lengthy randomized phase III trial was initiated in relapsed-refractory metastatic melanoma patients to determine overall survival. Their antibody, ipilimumab, was given to one arm, antibody plus a short peptide (gp100) for a melanoma differentiation antigen was given to a second arm and the control arm received the short peptide alone. Although the Kaplan–Meier plots of survival were inseparable for the first few months, a twofold survival benefit was detected at 12–15 months in both antibody arms, which was still durable after 2.5 years and included a complete response in some patients⁶⁰.

In a second BMS/Medarex randomized trial involving 502 patients with previously untreated metastatic melanoma, the addition of ipilimumab to standard dacarbazine therapy was shown to improve overall survival compared with dacarbazine alone (11.2 months versus 9.1 months). Furthermore, the combination treatment significantly increased the proportion of surviving patients with at least 3 years of follow-up (20.8% versus 12.2%)⁶¹. Although only a relatively small fraction of patients derived clinical benefit, these studies clearly establish ipilimumab as an active agent, which offers patients, who would normally be at the terminal stage of this disease, clinically meaningful benefits and the possibility of long-term survival. Furthermore, the results validate the idea that activating the T-cell compartment can, on its own, provide a significant therapeutic effect.

The use of ipilimumab does present some clinical and scientific challenges. First is the significant rate of on-target toxicities observed. Up to 23% of the ipilimumab-treated patients developed serious (grade 3–4) adverse events including colitis and hypophysitis due to induced inflammation (possibly autoimmune in nature), and in conjunction with

dacarbazine, approximately 20% showed significant elevations in liver function tests. However, toxicity does not accurately predict positive therapeutic outcome, indicating that many patients will experience inflammatory pathology without benefiting from an antitumour effect. These toxicities might be expected given that the removal of *CTLA4* from mice leads to virulent inflammatory disease, as mentioned earlier. The different spectrum of toxicities with ipilimumab compared with standard cancer treatments means that practising oncologists will need to acquire additional expertise in the management of inflammatory disorders.

A second clinical challenge with ipilimumab relates to the kinetics of the antitumour response. In contrast to conventional cytotoxic therapies that may trigger rapid tumour shrinkage due to direct killing of cancer cells, the stimulation of T-cell response with ipilimumab may take several months to occur. Tumours may increase in size during this period, and some component of this growth may be a result of the evolving inflammatory reaction. Indeed, as many as 10% of patients treated with ipilimumab, who were scored with progressive disease using the modified WHO (World Health Organization) criteria for tumour size, were shown to achieve disease stabilization and prolonged survival^{44,60}. This unusual pattern of treatment response has led to the proposal of new immune-related criteria that may aid clinical decision making regarding continuation of therapy⁶² (Box 3).

The rationale for ipilimumab monotherapy is that its use so far assumes that tumour-protective T cells exist in the patient before therapy, and that these cells will exert antitumour activity if *CTLA4* is blocked. The previously mentioned clinical studies were carried out without concomitant effective immunization. An uncoupled peptide to the melanoma differentiation antigen epitope was included in some arms, but dosed without adjuvant or a dendritic cell maturation agent. Although it would be expected for *CTLA4* to be induced in tumour-reactive T cells only after immunization, the clinical response observed in the absence of prior exogenous vaccination indicates that tumour-reactive TILs expressing *CTLA4* are responsive to checkpoint blockade

BOX 3

Clinical assessment of immunotherapy

Oncologists traditionally evaluate the activity of cancer therapies through measurements of tumour area or volume. These standard metrics include the RECIST and modified WHO criteria. Clinical responses to cytotoxic treatments, such as chemotherapy, radiation therapy and some targeted agents usually occur quickly (within a few weeks to months) because their presumed mechanism of action involves a direct effect on tumour cells. Moreover, these treatments generally result in a reduction in tumour size because cancer cells undergo apoptosis or other modes of programmed cell death. Although tumour regression indicates the therapy is beneficial, this may not always translate into improvements in survival due to the potential emergence of lethal drug-resistant cells. Immunotherapy-induced tumour destruction, in contrast, may be delayed or even preceded by a period of apparent tumour growth. In clinical trials of ipilimumab, 10–20% of patients showed an increase in tumour size when evaluated 3 months after starting treatment, but subsequently achieved prolonged tumour control or regression without any additional intervention. These patients demonstrated long-term survival comparable with patients who had more rapid tumour regression. The mechanisms underlying the delayed response are not yet well understood, but might include the effects of immune infiltrates in tumours or just the long period of time required to generate sufficient T cells to accomplish tumour killing. This distinctive biology has led to the proposal of immune-related response criteria⁴⁰, which allow for greater flexibility in following the increase in tumour size during immunotherapy before declaring treatment failure.

and acquire effective tumour-rejecting functions. The notion of exogenous versus endogenous vaccination is discussed later.

Despite these limitations, ipilimumab provides realistic hope for melanoma patients, particularly those with late stage disease who otherwise had little chance of survival. More broadly, it provides clear clinical validation for cancer immunotherapy in general. The results will also intensify the search for predictive biomarkers for positive responders. Other applications of ipilimumab are already being vigorously pursued, and the door has been opened for the development and investigation of a host of other potential immunotherapeutic strategies, some of which may prove safer and more effective than targeting *CTLA4*. At long last, there is the prospect of combining this treatment with other immunotherapeutic regimens, such as effective vaccination, which, arguably, should have been considered much earlier in the clinical study history of anti-*CTLA4*.

The next generation T-cell immunomodulators

The success of anti-*CTLA4* in melanoma should create interest in evaluating other antibodies that can be used to activate T-cell responses. There are a number of known receptors that could serve as targets for agonist antibodies, including 4-1BB, OX40, GITR, CD27 and CD28 (Fig. 3). The latter, however, introduces a cautionary note owing to an early clinical trial of an agonist anti-CD28 (TGN1412) in which severe toxicities and even death resulted from unexpected cytokine release⁶³. These serious events emphasize the power of the immune system and the need for extreme care and a conservative trial design when using any immune activator. The use of agents that clear more rapidly from the circulation than intact IgGs may help mitigate the potential for such toxicities, or at least enable the more rapid removal of the inducing drug. The same consideration may apply to anti-*CTLA4* therapy, where alternative dosing strategies may serve to increase its therapeutic index.

LAG-3 is another T-cell receptor that, like *CTLA4*, is largely suppressive. Not as well studied as *CTLA4*, LAG-3 appears similar in that it acts to limit the activity of CD4⁺ and CD8⁺ T cells, and augment the activity of T_{reg} cells^{64,65}. In the same way as *CTLA4*, there is also a significant intracellular pool of LAG-3 (ref. 66). However, the functional consequences of its deletion are far less dramatic because it may work alongside other regulatory molecules (for example, PD-L1)⁶⁷. This situation suggests, however, that antagonizing LAG-3 may provide an alternative to antagonizing *CTLA4*, and perhaps have a better safety profile.

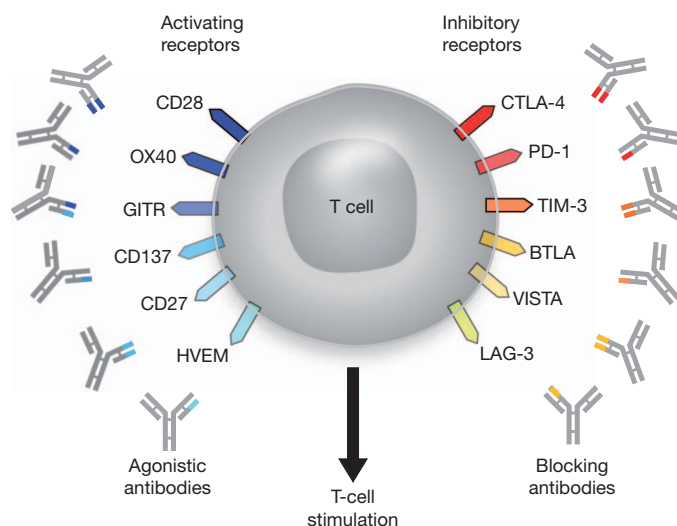


Figure 3 | T cell targets for immunoregulatory antibody therapy. In addition to specific antigen recognition through the TCR, T-cell activation is regulated through a balance of positive and negative signals provided by co-stimulatory receptors. These surface proteins are typically members of either the TNF receptor or B7 superfamilies. Agonistic antibodies directed against activating co-stimulatory molecules and blocking antibodies against negative co-stimulatory molecules may enhance T-cell stimulation to promote tumour destruction.

Another attractive approach, which is beginning to receive some clinical validation, is targeting immunosuppression in the tumour bed. Even if a vaccine or T-cell modulation therapy is successful, the ability of tumours to counteract immune effectors may limit the clinical benefit. Of current clinical interest is the PD-1/PD-L1(-L2) axis. PD-1 is expressed by T cells, particularly activated T cells, and binds to its ligands PD-L1/L2 that can be expressed by potential target cells, thereby rendering the T cell unresponsive or 'exhausted'⁶⁸. This axis is well-characterized as limiting the T-cell response to chronic virus infection⁶⁹, but increasingly it is thought to have a role in limiting the immune response in cancer as well⁷⁰. A variety of tumours, including melanoma, ovarian, renal, hepatocellular and glioblastoma, have been found to express PD-L1 (and occasionally PD-L2). PD-L1 expression has been found to correlate with poor prognosis^{71,72}, and to increase on activation of the oncogenic PI3K pathway (for example, by *PTEN* deletion)¹⁸. Antibodies to PD-1 have now reached the clinic. In early phase I trials they have shown good activity in a variety of cancer types and, so far, have a toxicity profile that seems safer than ipilimumab⁷³. The reduced toxicity is consistent with the generally milder autoimmune phenotype seen in *PD-1*^{-/-} mice⁶⁸ compared with *Ctla4*^{-/-} mice. However, lung inflammation and cardiomyopathy observed in mice^{74,75} may prove to be a concern in the clinic. An important biological and clinical question is whether the effector function of tumour-reactive TILs co-expressing multiple inhibitory receptors can be fully recovered by targeting a single receptor, or whether combinatorial checkpoint blockade is required for sustained tumour protection.

Other potential approaches to T-cell immunosuppression include targeting T_{reg} cells for inactivation or depletion because, in many tumour beds, the infiltration of T_{reg} cells may act to oppose effector T-cell function as much as the presence of negative regulatory molecules on the surface of effector cells (for example, PD-1). Whereas no specific surface marker of T_{reg} cells has yet been identified, some proteins, such as GITR and OX40, may be transiently expressed, possibly enabling them as targets (although they are also expressed by activated effector T cells in humans)^{76,77}. Anti-CD25 antibodies may preferentially deplete T_{reg}, at least following short-term therapy, and may help increase the efficacy of active immunization⁷⁸. Finally, low-dose cyclophosphamide may preferentially target T_{reg} and allow for attenuation of T_{reg} in the context of immunization protocols^{79,80}.

As mentioned earlier, tumour beds can produce a number of soluble mediators that counteract T-cell function, such as IDO (indoleamine 2,3-dioxygenase), arginase and prostaglandin E2 (PGE2)⁸¹; to the extent that these agents can be diagnostically demonstrated as predictive biomarkers. Efforts to inhibit their activities with either small molecule or antibody antagonists might also be useful additions to a portfolio of cancer immunotherapeutics. Furthermore, as knowledge on the immunoregulatory function of tumour vasculature increases, therapeutic manoeuvres to 'normalize' the tumour vessels in combination with immunotherapy seem to be an interesting development. Indeed, VEGF blockade has resulted in increased T-cell homing to tumours⁸² and has enhanced the efficacy of immunotherapy in the mouse⁸³. Lastly, the importance of tumour-promoting inflammatory pathways, such as STAT3 and NF- κ B signalling, and cytokines including IL-6, IL-17, IL-23 and tumour-necrosis factor- α (TNF- α) has been demonstrated⁸⁴. Inhibition of these circuits might not only antagonize tumour progression, but also enhance the activity of immunotherapy.

Combination immunotherapy

The use of combinations of chemotherapeutic drugs has traditionally been a mainstay of oncology. Even in the relatively recent age of molecularly targeted therapies, the development of combinations is proving to be of benefit to broaden the response and to treat resistance. The difference is, in theory, combinations of targeted agents can be combined rationally, in a scientifically guided fashion. There is no reason to believe that the same will not hold true for immunotherapy.

In melanoma, there has been another recent success in the small molecule realm. For patients bearing the V600E activating mutation of *BRAF*, the Roche inhibitor vemurafenib has been shown to yield a marked response in more than 50% of patients⁸⁵. Yet, resistance develops rapidly (<1 year), creating the necessity for additional therapy. Although one probable effective approach is to add a second small molecule inhibitor to *MEK*, to prevent the tumour from activating a compensatory pathway, one might predict the tumour will also circumvent this strategy. Therefore, the prospect of generating long-lasting protective immunity during the remission period is intensely attractive and trials combining ipilimumab with vemurafenib are in the planning stages. Since vemurafenib-induced death of tumours can be expected to release endogenous tumour antigens, it is possible that the small molecule and immunotherapeutic approaches may synergize, with *BRAF* inhibition acting to help prime *de novo* T-cell responses that can then be facilitated by the anti-CTLA4. Success, however, makes the assumption that the *BRAF* inhibitor is not suppressive of immune responses under the conditions that the two agents would be combined in the clinic.

In this and all other instances where immunotherapeutics are to be combined with targeted or chemotherapeutic agents, it will be critical to assess the potential interactions of the combination(s). Small molecule inhibitors, and cytotoxic chemotherapy alike, may act on cells of the immune system to block dendritic cells or T-cell function, or modify the tumour or tumour microenvironment in a way that will antagonize the development of immunity. Alternatively, conditions may be found where such inhibitors or cytotoxic drugs may be immunostimulatory to T cells, as recently suggested for an IAP antagonist⁸⁶ or, surprisingly, for the mTORC1 inhibitor, rapamycin, which is normally used as an immunosuppressive agent in the transplant setting⁸⁷. Similarly, although conventional wisdom viewed the effects of chemotherapy as obligatorily deleterious to immune mechanisms, such effects are drug-, dose- and/or schedule-dependent. As the same agent may prove inhibitory, benign or even stimulatory depending on the stage of immune response being targeted and the dose/schedule being used^{88,89}, great care must be exercised when designing clinical protocols in case an efficacious drug is dismissed due to negative interactions or suboptimal dosing schedules.

The effects of conventional chemotherapeutics on the immune system may be more nuanced than previously believed. Evidence is emerging that tumour cells can die in multiple ways, with some forms of (apoptotic) death actually leading to the enhancement of an immune response to the tumour⁹⁰. So-called immunogenic cell death is characterized in part by the release of ATP and high mobility group protein B1, which could activate local infiltrating myeloid cells and dendritic cells by a purinergic receptor or TLR4, respectively. Cytotoxic agents that elicit this death fingerprint may have the ability to help induce antitumour immune responses and, therefore, be better candidates for combination therapy with immunologically active agents. Again, great care must be exercised to ensure such agents are used at doses and schedules that do not suppress effector CTLs.

Finally, the prospect of combining immunotherapeutic agents themselves must be considered. To do this rationally, one must return again to the various steps of the immune response that need to be addressed to generate anticancer immunity (Fig. 1), consider the step(s) under the control of each agent, and assess the potential for overlapping or synergistic toxicity that would decrease rather than increase therapeutic index. For example, combining anti-CTLA4 with anti-PD-1 makes sense biologically, as the two agents remove the brakes from T-cell activation at two distinct stages: proliferation (CTLA4) and effector function (PD-1). Yet, both might be expected to show similar adverse events, underscoring the need to define carefully the potential for serious toxicity.

Revisiting vaccine use

Agents that act at the effector stage (for example, anti-PD-1 or inhibitors of immunosuppression) can only act by re-energizing the pre-existing T cells. Agents that act at the proliferation or activation stage (for example,

ipilimumab) can probably enhance not only pre-existing responses but also *de novo* responses. Thus, either could work well in conjunction with a vaccine approach, which is the place where immunotherapy in cancer got its start. However, it is debatable whether endogenous or exogenous vaccines would be preferable for this purpose.

Exogenous vaccines involve introducing preselected antigens on antibody delivery vehicles targeted to dendritic cells, encoded in viral vectors, or administered as peptides or proteins in a suitable adjuvant and carrier. The approach has been highly effective in generating prophylaxis and protective immunity against infectious agents⁹. In fact, such vaccines represent some of medicine's greatest successes. Vaccines have been far less impressive in cancer, which could be a reflection of the use of poor platforms that fail to elicit optimal antigen processing or presentation by dendritic cells, suboptimal adjuvants, or the absence of co-administered agents that facilitate T-cell responses or overcome immunosuppression in the tumour bed. In addition, the antigens selected may have been poorly chosen, unable to generate protective T cells with high affinity T-cell receptors because the tolerance barrier was too high or unable to generate a sufficient quantity of T-cell receptor ligands (peptide–MHC class I complexes) at the tumour cell surface.

Endogenous vaccines involve mobilizing antigens from a patient's own tumour, *in situ*. This would have to be accomplished by inducing tumour cell death under conditions that favour the ability of endogenous dendritic cells to capture, process and present tumour-derived antigens. Although less controllable than the exogenous approach, endogenous vaccines have the distinct advantage of potentially allowing the presentation of the dozens, or hundreds, of mutations harboured by cancer cells⁵. If even only a fraction of these can form MHC-binding peptides that serve as T-cell epitopes, then as neo-antigens they would certainly not have been subject to suppression by central tolerance. Whether chemotherapy alone is sufficient to induce such *in situ* vaccination, however, is unknown. Because of the powerful immunosuppressive circuitries operating in the tumour microenvironment, additional signals are probably required to correctly polarize the tumour-infiltrating-antigen presenting cells, which can be now achieved by pathogen-derived signals or other dendritic cell maturation agents.

A recent attempt at endogenous vaccination involves the intratumoral injection of replication-conditional herpes simplex viruses engineered to secrete GM-CSF. These oncolytic viruses (OncoVex), which have entered phase III testing in melanoma, cause lysis of tumour cells on infection, when GM-CSF release may enhance dendritic cell function in the tumour microenvironment to stimulate responses against autologous cancer antigens⁹¹. A second attractive approach involves the combination of cytotoxic therapies in conjunction with agonistic antibodies to CD40. Whereas this strategy might be predicted to augment dendritic cell priming of antitumour T cells, an important role for macrophage-mediated killing was implicated instead⁹². Lastly, combinations of chemotherapy with TLR agonists could provide the right ingredients of tumour antigen release and tumour antigen presenting cell activation.

Perspectives

Despite its long, if not always distinguished, history, these are early days for the new science and clinical practice of cancer immunotherapy. Nevertheless, as an exciting therapeutic strategy that merits serious consideration by the biopharmaceutical industry and clinical oncology community alike, recent results have allowed cancer immunotherapy to finally come of age. This transition coincides with another development of the cancer immunology field. For much of the twentieth century, the focus was on cancer 'immunosurveillance' given the prominent view that the immune system has a homeostatic role in controlling cancer. When oncogenic or other mutations occurred, the immune system was, in theory, thought to respond, thereby preventing the development of a tumour, at least when the system was operating normally. This concept was refined by 'immunoediting', which recognizes the complex and dynamic cross-talk between the tumour and the host throughout all stages of disease development, including the possibility of tolerance⁹³.

In the twenty-first century, however, we believe emerging pre-clinical and clinical data has turned attention directly towards the primary role of 'tolerance and immune suppression'. This leaves aside the issue of whether the immune response has a homeostatic function in prevention. The cancer develops and escapes control either because the tumour cells are too similar to their normal counterparts and, therefore, do not elicit sufficient immunity, or because they are adept at inducing peripheral tolerance. Whereas ipilimumab might work either to induce a *de novo* response or enhance a pre-existing one, the clinical activity of anti-PD-1 antibodies seems to be more closely associated with overcoming tolerance than with generating an anticancer T-cell repertoire.

Cancer immunotherapy has come of age at just the right time. The advent of a cohort of inhibitors that target oncogenic pathways with ever greater specificity is starting to reveal significant and sometimes spectacular responses in several indications. Yet, even in diagnostically defined populations, these responses can be transient or require continued dosing. If such drug regimens can be matched to appropriate immunotherapies, activating a patient's immune system during a time of tumour reduction and remission may be the best way to ensure that responses are converted to a long-term and durable benefit. As well as the development of additional agents to prime and guide the immune response, appropriate pharmacodynamic biomarkers will have to be implemented to determine if a given immunotherapy is having the desired effect and appropriate diagnostics used to identify which strategy to apply to which patient. Finally, scientists, drug developers and oncologists will have to work together to implement new metrics for evaluating the effectiveness of immunotherapies. Their mechanism of action is so distinct, both mechanistically and temporally, compared with conventional cytotoxic drugs, that they cannot be expected to perform according to standards developed a generation ago, even though the result may ultimately be curative.

1. Dougan, M. & Dranoff, G. Immune therapy for cancer. *Annu. Rev. Immunol.* **27**, 83–117 (2009).
2. Hall, S. S. *A Commotion in the Blood: Life, Death, and the Immune System* (Henry Holt, 1997).
3. Sylvester, R. J. Bacillus Calmette-Guerin treatment of non-muscle invasive bladder cancer. *Int. J. Urol.* **18**, 113–120 (2011).
4. Boon, T., Coulie, P. G., Van den Eynde, B. J. & van der Bruggen, P. Human T cell responses against melanoma. *Annu. Rev. Immunol.* **24**, 175–208 (2006).
5. Segal, N. H. *et al.* Epitope landscape in breast and colorectal cancer. *Cancer Res.* **68**, 889–892 (2008).
6. Mellman, I. & Steinman, R. M. Dendritic cells: specialized and regulated antigen processing machines. *Cell* **106**, 255–258 (2001).
7. Trombetta, E. S. & Mellman, I. Cell biology of antigen processing *in vitro* and *in vivo*. *Annu. Rev. Immunol.* **23**, 975–1028 (2005).
8. Zitvogel, L. & Kroemer, G. Anticancer immunotherapy using adjuvants with direct cytotoxic effects. *J. Clin. Invest.* **119**, 2127–2130 (2009).
9. **This paper reviews the argument for certain forms of apoptotic cell death as enhancing antitumour immune responses.**
10. Palucka, K., Banchereau, J. & Mellman, I. Designing vaccines based on biology of human dendritic cell subsets. *Immunity* **33**, 464–478 (2010).
11. Darrasse-Jeze, G. *et al.* Feedback control of regulatory T cell homeostasis by dendritic cells *in vivo*. *J. Exp. Med.* **206**, 1853–1862 (2009).
12. Jiang, A. *et al.* Disruption of E-cadherin-mediated adhesion induces a functionally distinct pathway of dendritic cell maturation. *Immunity* **27**, 610–624 (2007).
13. Steinman, R. M., Turley, S., Mellman, I. & Inaba, K. The induction of tolerance by dendritic cells that have captured apoptotic cells. *J. Exp. Med.* **191**, 411–416 (2000).
14. Steinman, R. M., Hawiger, D. & Nussenzweig, M. C. Tolerogenic dendritic cells. *Annu. Rev. Immunol.* **21**, 685–711 (2003).
15. Curiel, T. J. *et al.* Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nature Med.* **10**, 942–949 (2004).
16. Kono, K. *et al.* CD4⁺CD25^{high} regulatory T cells increase with tumor stage in patients with gastric and esophageal cancers. *Cancer Immunol. Immunother.* **55**, 1064–1071 (2006).
17. Kooi, S. *et al.* HLA class I expression on human ovarian carcinoma cells correlates with T-cell infiltration *in vivo* and T-cell expansion *in vitro* in low concentrations of recombinant interleukin-2. *Cell. Immunol.* **174**, 116–128 (1996).
18. Hamanishi, J. *et al.* Programmed cell death 1 ligand 1 and tumor-infiltrating CD8⁺ T lymphocytes are prognostic factors of human ovarian cancer. *Proc. Natl Acad. Sci. USA* **104**, 3360–3365 (2007).
19. Parsa, A. T. *et al.* Loss of tumor suppressor PTEN function increases B7–H1 expression and immunoresistance in glioma. *Nature Med.* **13**, 84–88 (2006).
20. Mellor, A. L. & Munn, D. H. IDO expression by dendritic cells: tolerance and tryptophan catabolism. *Nature Rev. Immunol.* **4**, 762–774 (2004).

20. Munn, D. H. & Mellor, A. L. IDO and tolerance to tumors. *Trends Mol. Med.* **10**, 15–18 (2004).
21. Marigo, I., Dolcetti, L., Serafini, P., Zanovello, P. & Bronte, V. Tumor-induced tolerance and immune suppression by myeloid derived suppressor cells. *Immunol. Rev.* **222**, 162–179 (2008).
22. Ohta, A. *et al.* A2A adenosine receptor protects tumors from antitumor T cells. *Proc. Natl. Acad. Sci. USA* **103**, 13132–13137 (2006).
23. Facciabene, A. *et al.* Tumour hypoxia promotes tolerance and angiogenesis via CCL28 and T_{reg} cells. *Nature* **475**, 226–230 (2011).
24. Aggarwal, S. & Pittenger, M. F. Human mesenchymal stem cells modulate allogeneic immune cell responses. *Blood* **105**, 1815–1822 (2005).
25. Bouzin, C., Brouet, A., De Vries, J., Dewever, J. & Feron, O. Effects of vascular endothelial growth factor on the lymphocyte-endothelium interactions: identification of caveolin-1 and nitric oxide as control points of endothelial cell energy. *J. Immunol.* **178**, 1505–1511 (2007).
26. Buckanovich, R. J. *et al.* Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. *Nature Med.* **14**, 28–36 (2008).
27. Zhang, L. *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
28. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
29. Rosenberg, S. A., Yang, J. C. & Restifo, N. P. Cancer immunotherapy: moving beyond current vaccines. *Nature Med.* **10**, 909–915 (2004).
30. Schwartzentruber, D. J. *et al.* gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N. Engl. J. Med.* **364**, 2119–2127 (2011).
31. Kenter, G. G. *et al.* Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. *N. Engl. J. Med.* **361**, 1838–1847 (2009).
32. Leffers, N. *et al.* Immunization with a P53 synthetic long peptide vaccine induces P53-specific immune responses in ovarian cancer patients, a phase II trial. *Int. J. Cancer* **125**, 2104–2113 (2009).
33. Vansteenkiste, J. *et al.* Final results of a multi-center, double-blind, randomized, placebo-controlled phase II study to assess the efficacy of MAGE-A3 immunotherapeutic as adjuvant therapy in stage IB/II non-small cell lung cancer (NSCLC). *J. Clin. Oncol.* **25** (18S), 7554 (2007).
34. Freedman, A. *et al.* Placebo-controlled phase III trial of patient-specific immunotherapy with mitumotrimut-T and granulocyte-macrophage colony-stimulating factor after rituximab in patients with follicular lymphoma. *J. Clin. Oncol.* **27**, 3036–3043 (2009).
35. Kantoff, P. W. *et al.* Overall survival analysis of a phase II randomized controlled trial of a poxviral-based PSA-targeted immunotherapy in metastatic castration-resistant prostate cancer. *J. Clin. Oncol.* **28**, 1099–1105 (2010).
36. Neller, M. A., López, J. A. & Schmidt, C. W. Antigens for cancer immunotherapy. *Semin. Immunol.* **20**, 286–295 (2008).
37. Copier, J. & Dalglish, A. Whole-cell vaccines: a failure or a success waiting to happen? *Curr. Opin. Mol. Ther.* **12**, 14–20 (2010).
38. Schuler, G. Dendritic cells in cancer immunotherapy. *Eur. J. Immunol.* **40**, 2123–2130 (2010).
- The authors summarize the current state of the art use of autologous dendritic cell vaccines.**
39. Neller, M. A., Lopez, J. A. & Schmidt, C. W. Antigens for cancer immunotherapy. *Semin. Immunol.* **20**, 286–295 (2008).
40. Cheever, M. A. *et al.* Translational Research Working Group developmental pathway for immune response modifiers. *Clin. Cancer Res.* **14**, 5692–5699 (2008).
41. Rammensee, H. G., Weischen, T., Gouttefangeas, C. & Stevanovic, S. Towards patient-specific tumor antigen selection for vaccination. *Immunol. Rev.* **188**, 164–176 (2002).
42. Small, E. J. *et al.* Immunotherapy of hormone-refractory prostate cancer with antigen-loaded dendritic cells. *J. Clin. Oncol.* **18**, 3894–3903 (2000).
43. Kantoff, P. W. *et al.* Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N. Engl. J. Med.* **363**, 411–422 (2010).
- This paper describes the clinical trial results that led to the FDA approval of Provenge for prostate cancer.**
44. Wolchok, J. D. *et al.* Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin. Cancer Res.* **15**, 7412–7420 (2009).
- The authors summarize the argument for using a different set of criteria to evaluate the efficacy of immunotherapies given the mechanistic differences from conventional anticancer drugs.**
45. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
46. Hodi, F. S. *et al.* Immunologic and clinical effects of antibody blockade of cytotoxic T lymphocyte-associated antigen 4 in previously vaccinated cancer patients. *Proc. Natl. Acad. Sci. USA* **105**, 3005–3010 (2008).
47. Dudley, M. E. *et al.* Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science* **298**, 850–854 (2002).
48. Rosenberg, S. A., Restifo, N. P., Yang, J. C., Morgan, R. A. & Dudley, M. E. Adoptive cell transfer: a clinical path to effective cancer immunotherapy. *Nature Rev. Cancer* **8**, 299–308 (2008).
49. June, C. H. Principles of adoptive T cell cancer therapy. *J. Clin. Invest.* **117**, 1204–1212 (2007).
50. Robbins, P. F. *et al.* Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J. Clin. Oncol.* **29**, 917–924 (2011).
51. Kalos, M. *et al.* T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci. Transl. Med.* **3**, 95ra73 (2011).
52. Topp, M. S. *et al.* Targeted therapy with the T-cell-engaging antibody blinatumomab of chemotherapy-refractory minimal residual disease in B-lineage acute lymphoblastic leukemia patients results in high response rate and prolonged leukemia-free survival. *J. Clin. Oncol.* **29**, 2493–2498 (2011).
- This paper describes the clinical results for an intriguing approach to tumour targeting of T cells using bispecific antibodies.**
53. Chambers, C. A., Kuhns, M. S., Egen, J. G. & Allison, J. P. CTLA-4-mediated inhibition in regulation of T cell responses: mechanisms and manipulation in tumor immunotherapy. *Annu. Rev. Immunol.* **19**, 565–594 (2001).
- A recent review describing the current understanding of the mechanism of action of CTLA-4.**
54. Waterhouse, P. *et al.* Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* **270**, 985–988 (1995).
55. Tivol, E. A. *et al.* Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* **3**, 541–547 (1995).
56. Wing, K. *et al.* CTLA-4 control over Foxp3⁺ regulatory T cell function. *Science* **322**, 271–275 (2008).
57. Cipponi, A., Wieers, G., van Baren, N. & Coulie, P. G. Tumor-infiltrating lymphocytes: apparently good for melanoma patients. But why? *Cancer Immunol. Immunother.* **60**, 1153–1160 (2011).
58. Disis, M. L. *et al.* High-titer HER-2/neu protein-specific antibody can be detected in patients with early-stage breast cancer. *J. Clin. Oncol.* **15**, 3363–3367 (1997).
59. Novellino, L., Castelli, C. & Parmiani, G. A listing of human tumor antigens recognized by T cells: March 2004 update. *Cancer Immunol. Immunother.* **54**, 187–207 (2005).
60. Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
- This paper reports the pivotal clinical trial data that shortly thereafter led to the FDA approval of anti-CTLA4 for the treatment of metastatic melanoma.**
61. Robert, C. *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).
62. Hoos, A. *et al.* Improved endpoints for cancer immunotherapy trials. *J. Natl. Cancer Inst.* **102**, 1388–1397 (2010).
63. Suntharalingam, G. *et al.* Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N. Engl. J. Med.* **355**, 1018–1028 (2006).
64. Workman, C. J. & Vignali, D. A. Negative regulation of T cell homeostasis by lymphocyte activation gene-3 (CD223). *J. Immunol.* **174**, 688–695 (2005).
65. Huang, C. T. *et al.* Role of LAG-3 in regulatory T cells. *Immunity* **21**, 503–513 (2004).
66. Woo, S. R. *et al.* Differential subcellular localization of the regulatory T-cell protein LAG-3 and the coreceptor CD4. *Eur. J. Immunol.* **40**, 1768–1777 (2010).
67. Okazaki, T. *et al.* PD-1 and LAG-3 inhibitory co-receptors act synergistically to prevent autoimmunity in mice. *J. Exp. Med.* **208**, 395–407 (2011).
68. Keir, M. E., Butte, M. J., Freeman, G. J. & Sharpe, A. H. PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* **26**, 677–704 (2008).
- This article reviews the evidence for immunosuppression by the induction of T cell exhaustion via activation of PD-1.**
69. Watanabe, T., Bertoletti, A. & Tanoto, T. A. PD-1/PD-L1 pathway and T-cell exhaustion in chronic hepatitis virus infection. *J. Viral Hepat.* **17**, 453–458 (2010).
70. Sakuishi, K. *et al.* Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J. Exp. Med.* **207**, 2187–2194 (2010).
71. Gadiot, J. *et al.* Overall survival and PD-L1 expression in metastasized malignant melanoma. *Cancer* **117**, 2192–2201 (2011).
72. Gao, Q. *et al.* Overexpression of PD-L1 significantly associates with tumor aggressiveness and postoperative recurrence in human hepatocellular carcinoma. *Clin. Cancer Res.* **15**, 971–979 (2009).
73. Brahmer, J. R. *et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J. Clin. Oncol.* **28**, 3167–3175 (2010).
74. Nishimura, H., Nose, M., Hiai, H., Minato, N. & Honjo, T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* **11**, 141–151 (1999).
75. Nishimura, H. *et al.* Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. *Science* **291**, 319–322 (2001).
76. Ito, T. *et al.* OX40 ligand shuts down IL-10-producing regulatory T cells. *Proc. Natl. Acad. Sci. USA* **103**, 13138–13143 (2006).
77. Cohen, A. D. *et al.* Agonist anti-GITR monoclonal antibody induces melanoma tumor immunity in mice by altering regulatory T cell stability and intra-tumor accumulation. *PLoS ONE* **5**, e10436 (2010).
78. Golovina, T. N. & Vonderheide, R. H. Regulatory T cells: overcoming suppression of T-cell immunity. *Cancer J.* **16**, 342–347 (2010).
79. Emens, L. A. *et al.* Timed sequential treatment with cyclophosphamide, doxorubicin, and an allogeneic granulocyte-macrophage colony-stimulating factor-secreting breast tumor vaccine: a chemotherapy dose-ranging factorial study of safety and immune activation. *J. Clin. Oncol.* **27**, 5911–5918 (2009).
80. Ghiringhelli, F. *et al.* Metronomic cyclophosphamide regimen selectively depletes CD4⁺CD25⁺ regulatory T cells and restores T and NK effector functions in end stage cancer patients. *Cancer Immunol. Immunother.* **56**, 641–648 (2007).
81. Muller, A. J. & Scherle, P. A. Targeting the mechanisms of tumoral immune tolerance with small-molecule inhibitors. *Nature Rev. Cancer* **6**, 613–625 (2006).
82. Manning, E. A. *et al.* A vascular endothelial growth factor receptor-2 inhibitor enhances antitumor immunity through an immune-based mechanism. *Clin. Cancer Res.* **13**, 3951–3959 (2007).
83. Shrimali, R. K. *et al.* Antiangiogenic agents can increase lymphocyte infiltration into tumor and enhance the effectiveness of adoptive immunotherapy of cancer. *Cancer Res.* **70**, 6171–6180 (2010).

84. Wang, L. *et al.* IL-17 can promote tumor growth through an IL-6-Stat3 signaling pathway. *J. Exp. Med.* **206**, 1457–1464 (2009).
85. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
86. Dougan, M. *et al.* IAP inhibitors enhance co-stimulation to promote tumor immunity. *J. Exp. Med.* **207**, 2195–2206 (2010).
87. Ferrer, I. R. *et al.* Cutting edge: rapamycin augments pathogen-specific but not graft-reactive CD8⁺ T cell responses. *J. Immunol.* **185**, 2004–2008 (2010).
88. Machiels, J. P. *et al.* Cyclophosphamide, doxorubicin, and paclitaxel enhance the antitumor immune response of granulocyte/macrophage-colony stimulating factor-secreting whole-cell vaccines in HER-2/neu tolerized mice. *Cancer Res.* **61**, 3689–3697 (2001).
89. Alagkiozidis, I. *et al.* Time-dependent cytotoxic drugs selectively cooperate with IL-18 for cancer chemo-immunotherapy. *J. Transl. Med.* **9**, 77 (2011).
90. Obeid, M. *et al.* Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nature Med.* **13**, 54–61 (2006).
91. Senzer, N. N. *et al.* Phase II clinical trial of a granulocyte-macrophage colony-stimulating factor-encoding, second-generation oncolytic herpesvirus in patients with unresectable metastatic melanoma. *J. Clin. Oncol.* **27**, 5763–5771 (2009).
92. Beatty, G. L. *et al.* CD40 agonists alter tumor stroma and show efficacy against pancreatic carcinoma in mice and humans. *Science* **331**, 1612–1616 (2011).
Although it is unclear if activation of CD40 works by activating innate or adaptive immunity, the results reported in this paper are impressive.
93. Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570 (2011).
94. Nahta, R. & Esteva, F. J. Herceptin: mechanisms of action and resistance. *Cancer Lett.* **232**, 123–138 (2006).
95. Taylor, C. *et al.* Augmented HER-2 specific immunity during treatment with trastuzumab and chemotherapy. *Clin. Cancer Res.* **13**, 5133–5143 (2007).
96. Lewis Phillips, G. D. *et al.* Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res.* **68**, 9280–9290 (2008).
97. Porter, D. L., Levine, B. L., Kalos, M., Bagg, A. & June, C. H. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N. Engl. J. Med.* **365**, 725–733 (2011).
A recent and clinically impressive attempt at adoptive T-cell therapy using genetically altered, patient-derived T cells.
98. Gogas, H. *et al.* Prognostic significance of autoimmunity during treatment of melanoma with interferon. *N. Engl. J. Med.* **354**, 709–718 (2006).
99. Shin, J. Y., Yoon, I. H., Kim, J. S., Kim, B. & Park, C. G. Vascular endothelial growth factor-induced chemotaxis and IL-10 from T cells. *Cell. Immunol.* **256**, 72–78 (2009).
100. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nature Rev. Immunol.* **9**, 162–174 (2009).

Acknowledgements This article is dedicated to the memory of our mentor, friend and inspiration Ralph Steinman, whose scientific life was dedicated to advancing the field of immunology in general and cancer immunotherapy in particular (he died shortly before being awarded the Nobel prize).

Author Contributions I.M. prepared the first draft of the manuscript, which was then modified by G.D., and further modified by G.C.; all three authors worked on and approved the final version.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to I.M. (mellman.ira@gene.com).

DNA-binding factors shape the mouse methylome at distal regulatory regions

Michael B. Stadler^{1,2*}, Rabih Murr^{1*}, Lukas Burger^{1,2*}, Robert Ivanek¹, Florian Lienert^{1,3}, Anne Schöler^{1,2,3}, Christiane Wirbelauer¹, Edward J. Oakeley⁴, Dimos Gaidatzis^{1,2}, Vijay K. Tiwari¹ & Dirk Schübeler^{1,3}

Methylation of cytosines is an essential epigenetic modification in mammalian genomes, yet the rules that govern methylation patterns remain largely elusive. To gain insights into this process, we generated base-pair-resolution mouse methylomes in stem cells and neuronal progenitors. Advanced quantitative analysis identified low-methylated regions (LMRs) with an average methylation of 30%. These represent CpG-poor distal regulatory regions as evidenced by location, DNase I hypersensitivity, presence of enhancer chromatin marks and enhancer activity in reporter assays. LMRs are occupied by DNA-binding factors and their binding is necessary and sufficient to create LMRs. A comparison of neuronal and stem-cell methylomes confirms this dependency, as cell-type-specific LMRs are occupied by cell-type-specific transcription factors. This study provides methylome references for the mouse and shows that DNA-binding factors locally influence DNA methylation, enabling the identification of active regulatory regions.

Methylation of cytosines in metazoan genomes adds epigenetic information onto DNA without changing the genetic information. It is thought to locally influence the regulatory potential of DNA, leading to changes in gene expression during development and disease^{1,2}. Highly parallel sequencing enables to determine DNA methylation genome-wide at base-pair and single-molecule resolution in unprecedented breadth and detail^{3,4}. Using whole-genome bisulphite sequencing (BisSeq), we generated such methylomes in mouse pluripotent embryonic stem (ES) cells and in neuronal progenitors (NP) and focused our analysis on methylation patterns that require high resolution for their detection. Using an analytical approach that quantifies DNA methylation locally, we identified a novel epigenome feature defined by localized reduced levels of methylation. These low-methylated regions (LMRs) show an average methylation of 30% and occur distal to promoters with little overlap with CpG islands. Contrasting LMRs with maps of histone modifications, DNase I hypersensitivity and several DNA-binding factors revealed that they are distal regulatory regions. Using the insulator protein CTCF and the transcriptional repressor REST as paradigms we show that DNA-binding factors are necessary and sufficient to create LMRs. LMRs form dynamically during differentiation driven by cell-type-specific factors, revealing a continuous crosstalk between DNA-binding factors and local DNA methylation.

Features of the mouse methylome

To obtain base-pair-resolution methylomes, we performed bisulphite conversion of DNA from mouse ES cells and corresponding NP derived through an established *in vitro* differentiation system⁵⁻⁷. High-coverage Illumina sequencing resulted in 77.3 billion mappable methylome bases. In parallel, we performed genomic sequencing to fully reconstruct the genotype together with available sequence data⁸. Our results show that cytosine methylation lies mostly in the context of CpG dinucleotides, and we observe non-CG methylation at levels that correspond to previous observations in human ES cells³ (Supplementary Fig. 1). Similar to the human methylome, however, non-CG methylation is generally found in regions already containing CpG methylation, and because it appears as a rare event we focused our

further analysis on CpG methylation. A global analysis yielded the anticipated DNA methylation pattern for mammalian cells^{3,4,7,9}. CpG methylation is low at promoter regions, whereas genic and intergenic regions show high methylation levels (Supplementary Fig. 1). The methylation frequency of individual CpGs has a bimodal distribution: the majority of CpGs show a high percentage of methylation, whereas a smaller group of cytosines are unmethylated (Fig. 1a).

Identification of low-methylated regions

In addition to this expected binary behaviour of CpG methylation (either hypo- or hyper-methylated), we also noticed cytosines with intermediate, yet low levels of methylation in the range of 10–50% (Fig. 1a). More than 4% of all CpGs (almost 1 million in the mouse genome) show this characteristic. When plotting methylation along chromosomal regions, cytosines with such methylation are not randomly distributed but cluster locally (Fig. 1b). We thus developed a computational approach that segments methylomes into regions of differential methylation. The applied Hidden Markov Model (HMM) takes into account only DNA methylation and no genome information such as CpG density or functional annotations. This segmentation identified three distinct classes: fully methylated regions (FMRs), unmethylated regions (UMRs) and low-methylated regions (LMRs) (Fig. 1a–c). It correctly identifies unmethylated CpG islands as UMRs^{3,4}, whereas most of the genome is classified as FMR, consistent with the fact that the majority of CpGs are methylated. The segmentation further identified LMRs at positions where local dips are visible in the raw data profiles, further supporting that these have distinct and identifiable methylation patterns (Fig. 1a–c). Importantly, LMRs are generally not CpG islands as they have lower CpG content, are shorter and they further reside distal to transcriptional start sites (TSS; Fig. 1c–f and Supplementary Fig. 1).

LMRs are distal regulatory regions

Interestingly, LMRs show evolutionary conservation (Fig. 2a), suggesting that they are of functional relevance. Further, they contain DNase-I-hypersensitive sites (DHS), a unique chromatin state that

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. ²Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. ³Faculty of Science, University of Basel, 4056 Basel, Switzerland. ⁴Novartis Institutes for BioMedical Research, Biomarker Development, 4056 Basel, Switzerland.

*These authors contributed equally to this work.

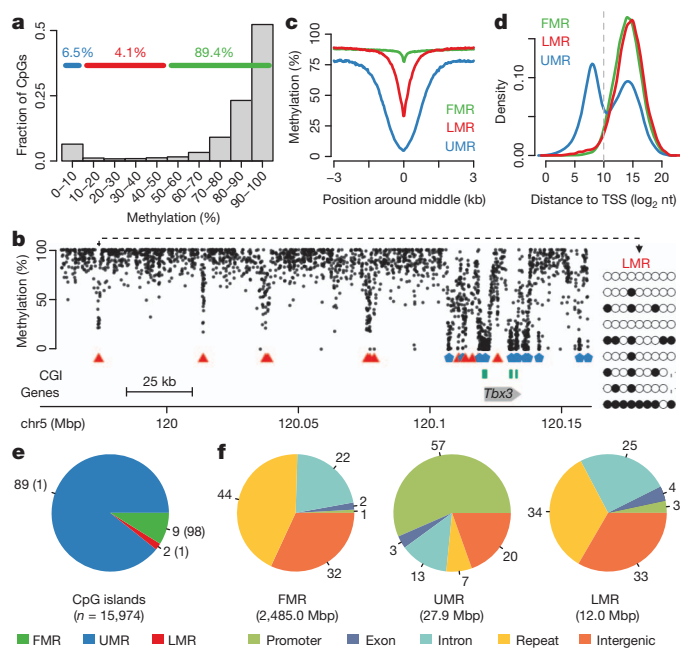


Figure 1 | Features of the mouse ES cell methylome. **a**, Distribution of CpG methylation frequency for all CpGs with at least tenfold coverage. Of all cytosines, 4.1% show intermediate methylation levels. **b**, Representative genomic region. Computational segmentation identifies UMRs (blue pentagons), LMRs (red triangles) and FMRs (unmarked). Each dot represents one CpG (CpG islands marked in green). Included is an independently verified LMR upstream of *Tbx3*. Mbp, million base pairs. **c**, Composite profile of CpG methylation for all three groups. kb, kilobases. **d**, Distances to TSS. **e**, **f**, Distribution of all three classes among genome features. **e**, A small percentage of LMRs overlap with CpG islands. Numbers indicate observed percentage of overlaps per group (expected percentage in parentheses). **f**, Distribution of the regions throughout the genome.

depends on DNA-binding factors^{10–12}. In fact, at least 80% of LMRs and 90% of UMRs overlap with DHS (Fig. 2 and Supplementary Fig. 2). LMRs are unlikely novel promoters as we find only weak signal for RNA polymerase II (Fig. 2 and Supplementary Fig. 3) and no RNA signal above what we observe at methylated regions even when using a strand-specific protocol that does not require polyadenylation (Supplementary Fig. 3). Next, we explored if LMRs could represent distal regulatory regions, such as enhancers. Indeed, LMRs are strongly enriched for chromatin features such as high H3K4 monomethylation (H3K4me1) signal relative to H3K4 trimethylation (H3K4me3) and the presence of p300 histone acetyltransferase, which are predictive features of enhancers¹³ (Fig. 2). This indicates that a subset of LMRs are enhancers that, in light of the absence of H3K27me3 and the presence of H3K27ac, are presumably active¹⁴ (Fig. 2b). Transgenic assays further show that individual LMRs increase the activity of a linked promoter and experimentally function as enhancers (Supplementary Fig. 4). We thus conclude that many LMRs, identified solely by their DNA methylation pattern, represent active regulatory regions.

To investigate LMR features further, we combined newly generated and published data sets for several DNA-binding factors and additional histone modifications (Supplementary Table 1, Fig. 2b and Supplementary Figs 5 and 6). LMRs and UMRs are depleted for the heterochromatic histone modification H3K9me2 in agreement with the absence of this mark at active chromatin⁶. Most DNA-binding factors show enrichment not only at UMRs, which are mostly promoters, but also at LMRs. Factors enriched at LMRs in stem cells include pluripotency transcription factors such as Nanog, Oct4 and Klf4, but also structural DNA-binding factors such as the insulator protein CTCF¹⁵ and members of the cohesin complex (Fig. 2b and

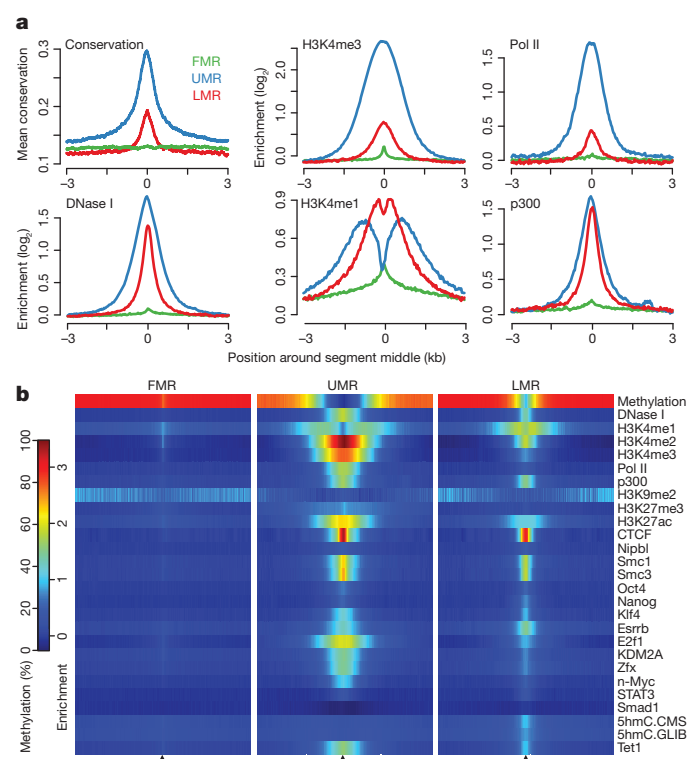


Figure 2 | General features of LMRs. Composite profiles 3 kb around segment midpoints. **a**, Evolutionary conservation based on multi-species alignments (upper left). Enrichment of DNase I tags (lower left). Chromatin features that predict enhancer function are enriched at LMRs (middle and right). **b**, Heat map of methylation levels, histone modifications and protein binding (H3K4me1 signal rescaled for visibility).

Supplementary Fig. 5), both of which bind promoters and distal regulatory regions¹⁶. Notably, not all factors occupy distal and proximal regulatory regions with equal preferences. Smad1 binds to neither LMRs nor UMRs, whereas some bind primarily at UMRs, such as KDM2A and Zfx, and others such as Nanog and Esrrb show higher enrichment at LMRs (Fig. 2b and Supplementary Fig. 5). In summary, several lines of evidence including genomic position, conservation, chromatin state, regulatory activity and transcription factor occupancy support the hypothesis that LMRs are indeed active distal regulatory regions.

Interestingly LMRs show a strong presence of 5-hydroxymethylcytosine (5hmC), consistent with recent reports of 5hmC presence at enhancer regions^{17–19}. One candidate protein responsible for catalysing 5hmC, Tet1 (refs 20, 21), is enriched at both UMRs and LMRs (Fig. 2b).

To ask if LMRs are also present in other mammals we performed HMM segmentation of a human stem cell methylome³, which also identifies LMRs with similar features, indicating that these are a general characteristic of mammalian methylomes (Supplementary Fig. 7).

Transcription factor binding creates LMRs

To determine how LMRs are formed, we investigated the DNA-binding protein CTCF, which binds to regulatory regions including promoters, enhancers and insulators^{22,23}. We determined the genome-wide binding of CTCF by chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Supplementary Fig. 8), revealing high occupancy at both UMRs and LMRs (Fig. 2b and Supplementary Fig. 5). A composite view of DNA methylation shows an average methylation of 20% at CTCF binding sites with increasing methylation adjacent to it (Supplementary Fig. 9), in line with a previous report in primates²⁴. If reduced methylation is a general feature of CTCF-occupied sites, inclusion of DNA methylation data should improve prediction of CTCF binding. Indeed, a binding model for CTCF that, in addition to the DNA

sequence, takes the DNA methylation state into account significantly increases the accuracy of prediction, illustrating that methylomes can improve transcription factor binding prediction (Fig. 3a). This however does not answer whether reduced methylation requires CTCF binding, a question that we addressed by a transgenic approach. We inserted reporter constructs with or without a CTCF-binding motif into the same genomic locus in mouse ES cells by a recombinase-based targeting approach²⁵ and tested the effect of CTCF binding on DNA methylation (Fig. 3b and Supplementary Fig. 9). Constructs without a CTCF site or with a mutated site became methylated when inserted, whereas introduction of a functional site led to CTCF binding and to locally reduced methylation (Fig. 3b and Supplementary Fig. 9). Thus CTCF binding seems necessary and sufficient to create an LMR. This predicts that endogenous differences in CTCF binding caused by single nucleotide polymorphisms (SNPs) should result in different methylation levels. Because the stem cells used in this study have a mixed genetic background between the C57BL/6 and 129S5 strains, we were able to identify regions that are heterozygous for SNPs within CTCF binding sites. An analysis of allele-specific CTCF binding and methylation confirms that point mutations that reduce CTCF binding lead to increased methylation at affected alleles (Fig. 3c and Supplementary Fig. 10). Together these experiments establish, for CTCF, that transcription factor binding is required for the formation of LMRs. To explore if CTCF binding can occur upstream of a change in DNA methylation, we inserted reporter constructs in an already methylated state generated *in vitro* before insertion. In the tested cases, the pre-existing methylation did not block CTCF binding and similarly resulted in reduced methylation (Fig. 3b

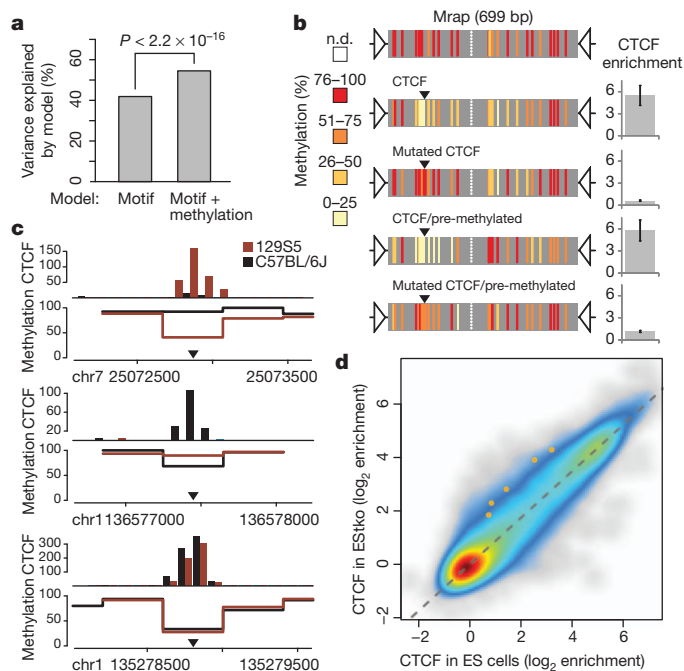


Figure 3 | DNA binding is necessary and sufficient for LMR formation. **a**, Including BisSeq data improves prediction of CTCF binding, measured as per cent of explained variance (100% = perfect prediction). **b**, Introduction of a CTCF motif (black triangle) into an otherwise methylated reporter construct leads to local demethylation. Insertion of a mutated site has no effect. Pre-methylation of the CTCF-motif-containing construct does not inhibit demethylation. Bar graphs show average CTCF enrichment from three experiments (error bars, s.d.). **c**, Allele-specific CTCF binding and DNA methylation for three loci. Two contain a heterozygous SNP (top and middle) and one serves as control (bottom). ChIP-seq and DNA methylation were quantified in 100-bp and 400-bp windows, respectively. **d**, CTCF binding in cells with or without DNA methylation. CTCF binding is globally unchanged in cells without methylation (ES_tko), but increases by approximately twofold at sites located in the H19 ICR (orange dots).

and Supplementary Fig. 9). We conclude that CTCF can bind to a pre-methylated CpG-poor target site, which in turn leads to localized demethylation. If this reflects the general binding model for CTCF, then its genomic binding sites should not be confined by DNA methylation. We tested this prediction by measuring CTCF binding in mouse ES cells that lack DNA methylation²⁶. In these cells, CTCF binding is indeed globally unaltered (Fig. 3d) with the exception of several paternally imprinted regions, such as the H19 ICR, where DNA methylation of selected CpG islands has previously been shown to block CTCF binding to one allele. At these CpG-rich regions we detect the expected twofold increase in CTCF binding in the absence of DNA methylation^{27,28} (Fig. 3d and Supplementary Fig. 10). Together our results indicate that CTCF binding within CpG-poor regions is generally not affected by the methylation status of its binding site, but rather that binding itself initiates local demethylation.

To ask whether the absence of a DNA-binding factor results in remethylation of endogenous LMRs, we analysed ES cells that lack the transcription factor REST. We chose *REST*^{-/-} ES cells because they are viable and pluripotent²⁹, whereas deleting CTCF is lethal³⁰. Indeed we observe increased DNA methylation in *REST*^{-/-} ES cells at LMRs that are occupied by REST in wild-type cells (Fig. 4 and Supplementary Fig. 11). Importantly, reintroduction of REST into *REST*^{-/-} cells leads to reformation of REST-dependent LMRs (Fig. 4 and Supplementary Fig. 11). Thus, also in the case of REST, DNA binding can occur in methylated CpG-poor regions, which in turn causes demethylation. These functional tests support a model where transcription factor binding is both sufficient and necessary for LMR creation. Although it remains to be tested whether the same hierarchy applies to other factors, this model predicts that LMRs should, at least in part, change during differentiation as a function of cell-type-specific DNA-binding factors.

Cell-type specificity of LMRs

To test this, we analysed the methylome of NP that were differentiated from ES cells^{6,7,25,31}. Similar to ES cells, the majority of the genome is methylated in NP whereas CpG islands remain mostly unmethylated (Supplementary Figs 12 and 13). Notably we do not observe partially methylated domains as described previously for human fibroblasts³ in ES cells nor NP. Segmentation of the NP methylome identifies LMRs with similar properties as those in ES cells (Fig. 5a and Supplementary Fig. 12). Chromosomal profiles of selected regions illustrate dynamic

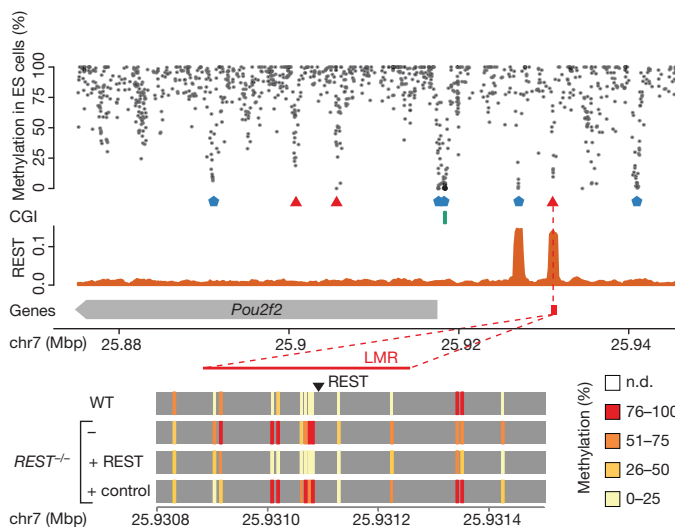


Figure 4 | REST is required for LMR formation at its binding sites. Bisulphite profile of a genomic region containing a REST binding site coinciding with an LMR. This LMR regains methylation in the absence of REST (*REST*^{-/-} ES cells), whereas reintroduction of REST re-establishes it.

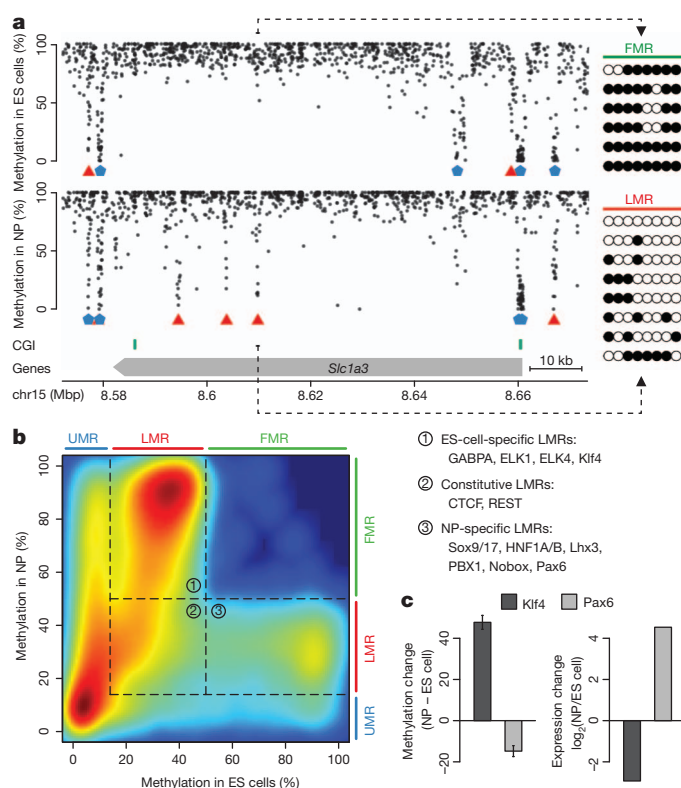


Figure 5 | Methylation dynamics during differentiation. **a**, Methylation of individual CpGs at an NP-specific LMR in a gene important for neuronal differentiation (*Slc1a3*). **b**, Comparison of DNA methylation in ES cells and NP of all CG-low regions (density increases from blue to red). LMRs were grouped into ES-cell-specific LMRs (box 1), constitutive LMRs (box 2) and NP-specific LMRs (box 3) and analysed for enriched DNA-binding motifs (indicated on the right). **c**, Methylation versus expression changes for Klf4 and Pax6. Klf4 binding sites in ES cells gain methylation in NP, whereas Pax6 binding sites lose methylation (top 200 sites; $P < 2.2 \times 10^{-16}$ for Klf4; $P = 7.4 \times 10^{-7}$ for Pax6; error bars, s.e.m.). Changes in methylation at binding sites are anti-correlated to expression changes of the respective factors (right graph).

local changes between ES cells and NP. In particular, new LMRs are formed adjacent to genes important for neuronal development (Fig. 5a and Supplementary Fig. 13). To quantify these changes, we grouped LMRs as conserved ($n = 8,748$), ES-cell-specific ($n = 22,184$) or NP-specific ($n = 3,773$) (Fig. 5b), which showed that only a minor fraction of LMRs occur in both cell states. UMRs on the other hand are mostly stable, in particular UMRs that are CpG-rich (Supplementary Fig. 12), in line with our previous study using low-resolution MeDIP-chip⁷. Dynamics of LMRs closely follow the dynamics of H3K4me1, a mark that is predictive for active enhancers and is strongly enriched at LMRs (Fig. 2b and Supplementary Fig. 14).

To ask how LMR dynamics are linked to gene regulation, we calculated expression changes for genes neighbouring dynamic and constitutive LMRs. This showed that LMR formation significantly correlates with increased expression of the nearest gene, whereas remethylation of LMRs coincides with reduced expression ($P = 2.3 \times 10^{-197}$, Supplementary Fig. 14). If LMR formation is indeed a function of DNA-binding factors, cell-type-specific LMRs should be enriched for DNA motifs of cell-type-specific DNA-binding proteins. Analysis of motif abundance in constitutive LMRs reveals over-representation of binding sites for factors present in both cell types, such as CTCF or REST ($P < 2.2 \times 10^{-16}$, Fig. 5b and Supplementary Fig. 15). In ES-cell-specific LMRs however, binding sites for factors related to pluripotency, such as Klf4 (ref. 32) and GABPA (ref. 33), are overrepresented ($P < 2.2 \times 10^{-16}$, Fig. 5b and Supplementary Fig. 15). Similarly, LMRs that are NP-specific are enriched for motifs of several neuronal

transcription factors, most notably Sox17 (ref. 34) and Pax6 (ref. 5) ($P < 2.2 \times 10^{-16}$, Fig. 5b and Supplementary Fig. 15). Next we compared the actual binding of the stem-cell-specific factor Klf4, for which binding sites were previously mapped in ES cells³⁵ and the NP-specific factor Pax6, for which we mapped binding sites by ChIP-chip in NP. In both cases, measured binding sites are less methylated in the cell type where the protein is present ($P < 2.2 \times 10^{-16}$, Fig. 5c and Supplementary Fig. 16). This finding fully agrees with our observation in *REST*^{-/-} cells (Fig. 4) and further links actual factor occupancy to LMR dynamics. To test the general utility of methylome profiles in reconstructing transcription factor activity, we generated a linear model that takes methylome data and sequence-predicted transcription factor binding sites as input and predicts individual contributions of the respective transcription factors for actual measured methylation differences as output. By using only the limited set of 126 DNA binding matrices from the core Jaspar database³⁶, this model explains almost 20% of the variation in observed methylation differences as a function of binding sites that occur within LMRs. Notably, the predicted activities of individual factors are in agreement with their actual expression changes (Supplementary Fig. 15). We thus conclude that methylation changes at CpG-poor regulatory regions are frequent during cellular differentiation and linked to expression changes of the nearest genes. Because these LMR dynamics are driven by transcription factor binding, methylomes can help to reconstruct distal regulation in a cell-type-specific fashion.

Discussion

This study provides base-pair resolution methylomes of the mouse as an important model organism. This resource was generated for pluripotent ES cells and for derived NP as a differentiated cell type. By implementation of an unsupervised segmentation approach that only considers DNA methylation, we identified LMRs as localized regions with reduced methylation. Importantly these regions do not resemble CpG islands, which are a different class of well-described hypomethylated regions in the genome³⁷ and that we identify as UMRs. LMRs are further flanked by fully methylated regions and are thus distinct from partial and dynamic methylation reported for the borders of CpG islands^{38,39}. LMRs localize within CpG-poor regions and show a residual level of methylation.

We identify LMRs as a discrete class of regulatory regions that are distal to promoters, which shows that reduced methylation at regulatory regions is a general phenomenon that extends beyond CpG islands. This is in agreement with several single-gene studies of methylation dynamics at localized distal regulatory regions^{24,40–45}. Our results are further compatible with the observation of CpG-poor regions with low methylation levels in Rhesus ES cells²⁴ and recent reports of reduced methylation outside of CpG islands^{38,46}.

We provide several lines of evidence that LMRs are a methylome feature that identifies distal regulatory regions: they are evolutionary conserved, are DNase I hypersensitive, show chromatin marks of enhancers, enhance expression of a linked promoter and are occupied by transcription factors. Importantly, we show that the human methylome has very similar LMR characteristics, suggesting that LMRs are generally present in mammalian epigenomes. Our data strongly suggest that LMRs result from transcription factor binding: (1) LMRs are highly dynamic during differentiation, as a subset is lost while new sites form. These dynamics can be explained directly by differential presence and activity of DNA-binding factors, highlighting that DNA methylation is more variable at LMRs than, for example, at CpG islands; (2) we provide evidence for REST and CTCF that LMR presence requires transcription factor binding; and (3) for both factors we show that binding within CpG-poor regions occurs before demethylation and LMR formation.

Our finding that the methylome is highly dynamic outside of CpG islands as a function of DNA-binding factors provides a framework to explain local methylation changes, many of which have been

correlated to changes in expression of neighbouring genes. While it remains to be determined for more factors if demethylation occurs before or after factor binding, we do show that binding of CTCF can occur on a CpG-poor but methylated transgenic template and that REST, when reintroduced into a *REST*^{-/-} background, leads to binding at methylated regions and coinciding demethylation. It is conceivable that factors such as CTCF and REST, both of which bind to a large binding motif with high specificity, mediate demethylation, which in turn enables binding of additional DNA- or chromatin-binding factors that are sensitive to DNA methylation. Regardless, our functional tests show that methylation changes can occur downstream of binding and thus can reflect ongoing activity of DNA-binding factors rather than regulating their binding.

The identified dynamics of DNA methylation provide an unexpected rich source of information. Indeed, we show that DNA modification knowledge alone allows for more precise identification of transcription factor binding sites and the unbiased discovery of active distal regulatory regions. This is not only relevant for our understanding of epigenome formation and function but has direct practical applications. Current methods to identify enhancers by ChIP¹³ or DNase I hypersensitivity⁴⁷ rely on sufficient quantity of fresh cell material, which can be limited or unavailable in primary samples. In contrast, methylome analysis does not require intact cells but only DNA in comparably small amounts. Furthermore, DNA methylation is chemically stable and can be analysed from any DNA material suitable for sequencing. Partly for that reason, BisSeq methylomes will be generated for primary samples within international cancer genome and human epigenome efforts^{48,49}. Our results add an additional dimension to such data sets and suggest that they can be exploited to reconstruct transcriptional regulation and to identify cell type and origin.

Our findings reveal that the mouse methylome is more dynamic during cellular differentiation than anticipated. These highly localized changes are linked to gene regulation with transcription factor binding likely to be the key responsible activity. Although further experiments are needed to address how both processes are linked, we note that the enrichment of TET1 and 5-hydroxymethylation at LMRs is compatible with a model in which the observed methylation dynamics could entail an active process.

METHODS SUMMARY

Biological materials and sequencing libraries. ES cells were cultured and differentiated as described⁵⁷. ChIP-seq libraries were prepared according to Illumina's instructions (Catalogue number IP-102-1001). Pax6 ChIP samples were amplified using a WGA2 kit (Sigma) and hybridized to a custom tiling microarray⁷ (Roche-NimbleGen). RNA-Seq libraries were generated according to Illumina's instructions using oligo-dT primers. RNA for strand-specific RNA-Seq was depleted from ribosomal RNA (rRNA) using Ribo-Zero rRNA removal kit (Epicentre) and libraries generated following the Illumina pre-release version of the Directional mRNA-Seq Library Preparation Guide. BisSeq library generation was adapted from Illumina Genomic DNA Sample Preparation Guide by ligating methylated adapters to fragmented genomic DNA followed by gel purification, sodium bisulphite conversion and seven cycles of PCR amplification. Libraries were sequenced using Illumina GA II and Illumina HiSeq 2000.

Read filtering, alignment, weighting and quantification of methylation levels. ChIP-seq and RNA-Seq: low-complexity reads were filtered out on the basis of dinucleotide entropy. Genome alignments were performed using bowtie⁵⁰. BisSeq: All C nucleotides in reads from bisulphite-converted samples were converted *in silico* to T nucleotides and then aligned to a similarly converted genome separately to each strand using bowtie⁵⁰. Per cent methylation was calculated for individual Cs as the ratio of the number of alignments with C (methylated), over the number of alignments with either C or T. Segmentation was performed using the R package RHmm with a three-state HMM corresponding to fully methylated, low-methylated and unmethylated CpGs. Segment coordinates in ES cells and NP are listed in Supplementary Tables 2 and 3. Heterozygous regions for allele-specific analyses were identified using a three-state HMM (C57BL/6J, 129S5 or heterozygous) on allelic ratios at known SNPs with high sequence coverage (Supplementary Table 4).

Received 28 June; accepted 15 November 2011.

Published online 14 December 2011.

- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
- Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
- Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
- Bibel, M., Richter, J., Lacroix, E. & Barde, Y. A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols* **2**, 1034–1043 (2007).
- Lienert, F. *et al.* Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet.* **7**, e1002090 (2011).
- Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
- Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
- Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42**, 1093–1100 (2010).
- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
- Sabo, P. J. *et al.* Genome-wide identification of DNase hypersensitive sites using active chromatin sequence libraries. *Proc. Natl Acad. Sci. USA* **101**, 4537–4542 (2004).
- Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
- Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
- Ohlsson, R., Bartkuhn, M. & Renkawitz, R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* **119**, 351–360 (2010).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
- Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
- Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
- Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
- Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).
- Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
- Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
- Cohen, N. M. *et al.* DNA methylation programming and reprogramming in primate embryonic stem cells. *Genome Res.* **19**, 2193–2201 (2009).
- Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genet.* **43**, 1091–1097 (2011).
- Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
- Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
- Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
- Jorgensen, H. F., Chen, Z. F., Merckenschlager, M. & Fisher, A. G. Is REST required for ESC pluripotency? *Nature* **457**, E4–E5 (2009).
- Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
- Bibel, M. *et al.* Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nature Neurosci.* **7**, 1003–1009 (2004).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Kinoshita, K. *et al.* GABP α regulates Oct-3/4 expression in mouse embryonic stem cells. *Biochem. Biophys. Res. Commun.* **353**, 686–691 (2007).
- Sohn, J. *et al.* Identification of Sox17 as a transcription factor that regulates oligodendrocyte development. *J. Neurosci.* **26**, 9722–9735 (2006).
- Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
- Bryne, J. C. *et al.* JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36**, D102–D106 (2008).

37. Illingworth, R. S. & Bird, A. P. CpG islands – ‘a rough guide’. *FEBS Lett.* **583**, 1713–1720 (2009).
38. Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **44**, 17–28 (2011).
39. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41**, 178–186 (2009).
40. Brunk, B. P., Goldhamer, D. J. & Emerson, C. P. Jr. Regulated demethylation of the myoD distal enhancer during skeletal myogenesis. *Dev. Biol.* **177**, 490–503 (1996).
41. Mareš, J. *et al.* Methylation changes in promoter and enhancer regions of the WT1 gene in Wilms’ tumours. *Cancer Lett.* **166**, 165–171 (2001).
42. Sharrard, R. M., Royds, J. A., Rogers, S. & Shorthouse, A. J. Patterns of methylation of the *c-myc* gene in human colorectal cancer progression. *Br. J. Cancer* **65**, 667–672 (1992).
43. Tagoh, H. *et al.* Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells *in vitro*. *Blood* **103**, 2950–2955 (2004).
44. Thomassin, H., Flavin, M., Espinas, M. L. & Grange, T. Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J.* **20**, 1974–1983 (2001).
45. Groudine, M. & Conkin, K. F. Chromatin structure and *de novo* methylation of sperm DNA: implications for activation of the paternal genome *Science* **228**, 1061–1068 (1985).
46. Colaneri, A. *et al.* Expanded methyl-sensitive cut counting reveals hypomethylation as an epigenetic state that highlights functional sequences of the genome. *Proc. Natl Acad. Sci. USA* **108**, 9715–9720 (2011).
47. Boyle, A. P. & Furey, T. S. High-resolution mapping studies of chromatin and gene regulatory elements. *Epigenomics* **1**, 319–329 (2009).
48. Abbott, A. Europe to map the human epigenome. *Nature* **477**, 518 (2011).
49. Satterlee, J. S., Schubeler, D. & Ng, H. H. Tackling the epigenome: challenges and opportunities for collaboration. *Nature Biotechnol.* **28**, 1039–1044 (2010).
50. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank A. Fernandez, C. Kohler, V. Petitjean and F. Staedtler (Novartis) and I. Nissen and C. Beisel (ETH-BSSE) for performing next generation sequencing experiments. R. Lister and J. Ecker for technical advice on BisSeq library generation. D. Schmitz for the pcDNA6-IRES-Blasticidin vector. D. Anderson for the REST antibody. H. Jørgensen for REST cds. L. Hoerner for help in Sanger bisulphite sequencing. M. Lorincz, N. Thomä and members of the Schübeler laboratory for feedback on the manuscript. R.M. is supported by an EMBO long-term postdoctoral fellowship. V.K.T. is supported by a Marie Curie International Incoming fellowship and an EMBO long-term postdoctoral fellowship. Research in the laboratory of D.S. is supported by the Novartis Research Foundation, the European Union (NoE “EpiGeneSys” FP7-HEALTH-2010-257082), the European Research Council (ERC EpiGePlas), the SNF Sinergia program and the Swiss initiative in Systems Biology (Cell Plasticity).

Author Contributions Experiments were designed by R.M., F.L., A.S., V.K.T., E.J.O. and D.S. BisSeq, RNA-Seq and ChIP-seq experiments were conducted by R.M., A.S. and V.K.T. ChIP-seq data analysis was performed by M.B.S. and L.B. BS-PCR validation was performed by R.M., F.L. and C.W. Sequencing data processing was performed by D.G. and M.B.S. LMRs were first noticed by D.G. Bioinformatic and statistical analyses were conducted by M.B.S., L.B. and R.I. The manuscript was prepared by R.M., M.B.S., L.B. and D.S.

Author Information Data sets generated for this study are available from GEO under accession GSE30206. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.S. (dirk@fmi.ch).

A compact system of small planets around a former red-giant star

S. Charpinet^{1,2}, G. Fontaine³, P. Brassard³, E. M. Green⁴, V. Van Grootel^{5,6}, S. K. Randall⁷, R. Silvotti⁸, A. S. Baran^{9,10}, R. H. Østensen¹¹, S. D. Kawaler¹⁰ & J. H. Telting¹²

Planets that orbit their parent star at less than about one astronomical unit (1 AU is the Earth–Sun distance) are expected to be engulfed when the star becomes a red giant¹. Previous observations have revealed the existence of post-red-giant host stars with giant planets^{2–4} orbiting as close as 0.116 AU or with brown dwarf companions^{5,6} in tight orbits, showing that these bodies can survive engulfment. What has remained unclear is whether planets can be dragged deeper into the red-giant envelope without being disrupted and whether the evolution of the parent star itself could be affected^{7–9}. Here we report the presence of two nearly Earth-sized bodies orbiting the post-red-giant, hot B subdwarf star KIC 05807616 at distances of 0.0060 and 0.0076 AU, with orbital periods of 5.7625 and 8.2293 hours, respectively. These bodies probably survived deep immersion in the former red-giant envelope. They may be the dense cores of evaporated giant planets that were transported closer to the star during the engulfment and triggered the mass loss necessary for the formation of the hot B subdwarf, which might also explain how some stars of this type did not form in binary systems.

KIC 05807616 (also known as KPD 1943+4058) is a seemingly isolated pulsating hot B subdwarf (sdB) star that has been monitored by the Kepler satellite primarily for the study of its oscillations^{10–12}. It is at an evolved stage of thermonuclear fusion of helium in its core, and belongs to the so-called extreme horizontal branch. This star shows a rich pulsation spectrum mostly composed of gravity (g-) modes^{13,14}. Its main structural parameters are well determined through asteroseismic means based on one month of Kepler exploratory data (Table 1)¹⁵. We focus here on additional Kepler time-series photometry obtained for this star after the exploratory phase.

The analysis of these data revealed many frequencies, most being associated with stellar oscillations (Fig. 1 and Supplementary Information section A1). However, we also found two very weak periodic modulations in the low-frequency range whose nature is most intriguing. The timescales involved for these variations are 5.7625 ± 0.0001 h (F_1 , with an amplitude of 52 ± 6 parts per million, p.p.m.) and 8.2293 ± 0.0003 h (F_2 , with an amplitude of ~ 47 p.p.m.; see Supplementary Information sections A2 and E). The phase folded curves show that these variations repeat at a coherent phase throughout the entire light curve (Fig. 1b and c). We determined that contamination from a nearby star, stellar pulsations, or rotational modulations (for example, through surface spots) cannot account for these variations, leaving orbital modulations as the most plausible interpretation (Fig. 1 and Supplementary Information sections B and C). Compact binary systems with an sdB star as the primary and having comparable orbital periods are indeed not atypical. However, in the present case, the very subtle modulations and the suggested presence of more than one companion raise the question of the

eventual substellar nature of these objects. In this context, the period ratio $F_2/F_1 \approx 1.43$, which is close to a 3:2 resonance, is also intriguing.

We explored this possibility by evaluating the properties (in particular the radius) that these objects would need to have in order to produce the observed variations in the light curve. There are two main sources of modulation involving small bodies. The first is reflection by

Table 1 | Derived parameters of KIC 05807616 and its two planet candidates

Stellar parameter ¹⁵	KIC 05807616	
Effective temperature, T_{eff} (K)	$27,730 \pm 270$	
Surface gravity, $\log[g \text{ (c.g.s.)}]$	5.52 ± 0.03	
Mass, M_* (M_{\odot})	0.496 ± 0.002	
Radius, R_* (R_{\odot})	0.203 ± 0.007	
Mean density, ρ_* (g cm^{-3})	84.1 ± 2.9	
Age after red-giant stage, A (Myr)	18.4 ± 1.0	
Bolometric luminosity, L (L_{\odot})	22.9 ± 3.1	
Apparent Johnson V-band magnitude, V	14.87 ± 0.02	
Distance from Earth, d (pc)	$1,180 \pm 95$	
Planetary parameter	Planet candidate 1 KOI 55.01	Planet candidate 2 KOI 55.02
Assumed Bond albedo*, α_j	0.10	0.10
Assumed temperature contrast†, β_j	0.2	0.2
Assumed inclination angle‡, i (degrees)	65	65
Assumed mean density§, ρ_j (g cm^{-3})	5.515	5.515
Orbital period, P_j (h)	5.7625 ± 0.0001	8.2293 ± 0.0003
Modulation amplitude, A_j (p.p.m.)	52 ± 6	~ 47
Orbit radius , a_j	$8.9698 \times 10^{10} \text{ cm}$ $1.290 R_{\odot}$ 0.0060 AU	$1.13749 \times 10^{11} \text{ cm}$ $1.636 R_{\odot}$ 0.0076 AU
Roche limit¶, d_R	0.0029 AU	
Mean temperature: day side#, T_j (K)	9,115	8,094
Mean temperature: night side#, T_j (dark) (K)	1,823	1,619
Planet radius☆, R_j (R_{\oplus})	0.759	0.867
Planet mass**, m_j (M_{\oplus})	0.440	0.655
Host star projected radial velocity††, v_j (m s^{-1})	0.65	0.86

Values shown assume the most probable configuration for this system. A third body (KOI 55.03) may be present between the two well secured detections (Fig. 1). If confirmed with more observations and using the same assumptions, its distance from the star would be 0.0065 AU, its radius $0.605 R_{\oplus}$, its mass $0.222 M_{\oplus}$, and its estimated temperature on the day (dark) side 8,735 K (1,747 K). Subscript @ indicates Earth value.

* We use a low albedo value representative of hot Jupiters. This value is also close to Mercury's albedo. † $\beta = T_j(\text{dark})/T_j$ is a parametrization of the thermal imbalance on the planet surface. On Mercury, $\beta \approx 0.2$.

‡ This inclination angle is suggested by the amplitude distribution of pulsation modes split by rotation.

§ Considering the derived radii, we assume rocky worlds with the Earth's density as a representative value.

|| Obtained from Kepler's third law (equation (D15), Supplementary information section D).

¶ Distance from the star below which tidal disruption occurs (assuming a rigid body), $d_R \approx R_*(2\rho_*/\rho_j)^{1/3}$.

Obtained assuming radiative equilibrium (equation (D5), Supplementary information section D).

☆ Derived from equation (D14) in Supplementary information section D.

** Derived from equation (D16) in Supplementary information section D.

†† Derived from equation (D17) in Supplementary information section D.

¹Université de Toulouse, UPS-OMP, IRAP, F-31400 Toulouse, France. ²CNRS, IRAP, 14 avenue Edouard Belin, F-31400 Toulouse, France. ³Département de Physique, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal QC H3C 3J7, Canada. ⁴Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. ⁵Institut d'Astrophysique et de Géophysique, Université de Liège, 17 Allée du 6 Août, B-4000 Liège, Belgium. ⁶FNRS, rue d'Egmont 5, B-1000 Bruxelles, Belgium. ⁷ESO, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, Germany. ⁸INAF-Osservatorio Astronomico di Torino, Strada dell'Osservatorio 20, 10025 Pino Torinese, Italy. ⁹Mt Suhora Observatory, Cracow Pedagogical University, ulica Podchorążych 2, 30-084 Krakow, Poland. ¹⁰Iowa State University, Department of Physics and Astronomy, 12 Physics Hall, Ames, Iowa 50011, USA. ¹¹Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium. ¹²Nordic Optical Telescope, Apartado 474, 38700 Santa Cruz de La Palma, Spain.

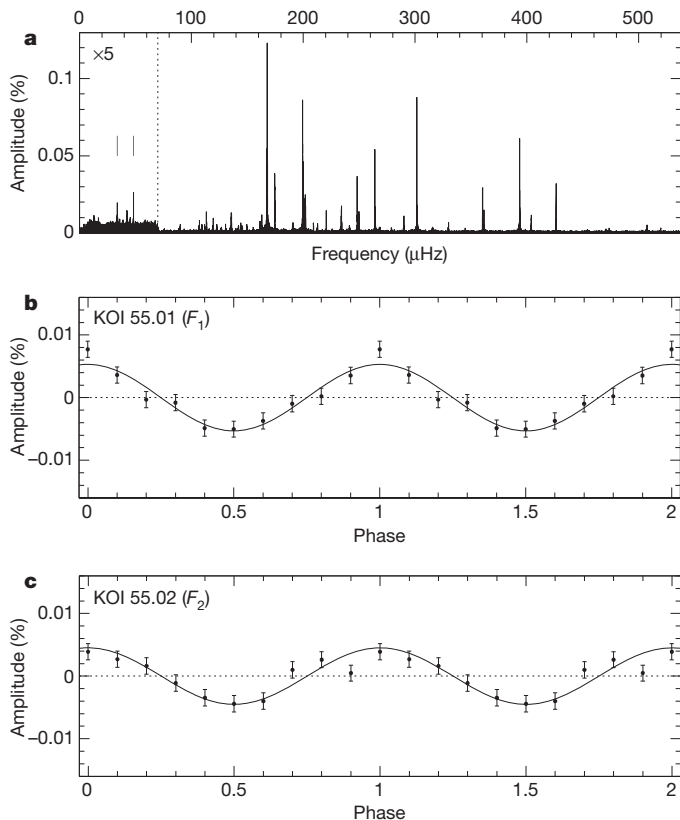


Figure 1 | Brightness variations detected in KIC 05807616. **a**, Amplitude spectrum of the light modulation; **b**, **c**, phase curves for the two planet candidates, KOI 55.01 (**b**) and KOI 55.02 (**c**). The combined Q2 + Q5–Q8 Kepler data (Qx refers to a 3-month observation quarter; see Supplementary Information section A1) have been used (that is, 14 months of monitoring spanning 21 months in total). In **a**, numerous peaks rising up to $\sim 0.15\%$ of the mean brightness of the star are due to g-mode pulsations (Supplementary Information section A1). Two weak variations indicated by vertical lines in the low frequency domain (shown with an amplitude expansion factor of 5) are also present, well disconnected from the g-modes. The frequency F_1 at $48.204 \pm 0.001 \mu\text{Hz}$ ($5.7625 \pm 0.0001 \text{ h}$) has a stable amplitude ($52 \pm 6 \text{ p.p.m.}$) and phase throughout the period of monitoring. F_2 at $33.755 \pm 0.001 \mu\text{Hz}$ ($8.2293 \pm 0.0003 \text{ h}$) is the dominant component of a more complex structure possibly caused by a frequency modulation (Supplementary Information sections A1 and E). After evaluating alternative interpretations (Supplementary Information sections B and C), we find that these structures are most likely to be the signatures of two bodies of substellar nature closely orbiting the star. We note that an even weaker signal between F_1 and F_2 (F_3 , $\sim 29 \text{ p.p.m.}$ at $42.4299 \mu\text{Hz}$ or 6.5467 h) could indicate the presence of a third body (KOI 55.03). In the following, we however concentrate on the two well secured detections. Panels **b** and **c** illustrate the light curve phase folded (in 10 phase bins) on the orbital period of KOI 55.01 (F_1) and KOI 55.02 (F_2), respectively (error bars, s.e.m.). These curves are obtained after ‘prewhitening’ the dominant g-mode pulsations. A small brightness excess occurring at phase zero in **b** is suggested. It may be typical of an opposition surge, but it is only a 1.9σ detection and more Kepler data will be needed to clearly establish its existence.

the surface of each planet of the light emitted by the star; such reflection would be modulated by the orbital motions, as we see the planets in different phases of illumination. The second source is a possible thermal imbalance between the heated day-side and cooler night-side hemisphere of each planet. Assuming the rotation of each planet is tidally synchronized to its orbital motion (a very likely situation for a compact system like KIC 05807616), we would see a thermal emission that was modulated along their orbits¹⁶. Both effects depend on unknown properties of the orbiting bodies. These are the amount of light reflected and absorbed at their surface (characterized by the Bond albedo, α) and how the absorbed heat is redistributed before being re-emitted. For this latter effect, we adopted a simple approach involving

average temperatures on the day and night sides, assuming radiative equilibrium and black-body re-emission. Heat redistribution is then characterized by a parameter, β , being the ratio of the mean temperature in the dark hemisphere to the mean temperature in the heated hemisphere. The magnitude of these variations also depends on the inclination angle of the system relative to the observer.

We calculated the planets’ properties as functions of the parameters mentioned above and explored all plausible situations (Supplementary Information section D). Our findings from this thorough analysis are as follows: (1) in all configurations except for very small inclination angles ($i < 20^\circ$), the derived radii for the orbiting objects are in the planetary range. (2) If efficient heat redistribution occurs ($\beta > 0.90$), the objects would be in the giant planet range of sizes, similar to Neptune or Jupiter in size. However, in most situations (when $\beta < 0.80$), the day/night temperature contrast effect dominates, and nearly Earth-sized planets are predicted. As this system is extremely compact and orbits a very hot star, the temperature contrast between the two hemispheres should be large, and therefore the value of β small, and the small-planet solution is preferred. (3) The assumed albedo has a limited influence on the above results and does not qualitatively change the conclusions. A low albedo, $\alpha \approx 0.10$, is generally observed for hot gaseous planets^{17–19} and Mercury, in our Solar System, has a comparable albedo (but, admittedly, it may not be representative of the hottest telluric worlds). The last unknown is the inclination angle of the system, and its effect is illustrated in Fig. 2. At very small inclinations ($i < 3^\circ$), the derived radius could still be in the giant-planet range. However, the pulsations detected in KIC 05807616 provide valuable indications that small angles ($i < 20^\circ$) are improbable, suggesting instead that $i \approx 65^\circ$. This additional constraint relies on the reasonable assumption that the orbits of the planets are most probably coplanar with the star’s equatorial plane (see Supplementary Information sections A3 and A4).

On the basis of these considerations, we have derived the parameters for the two planet candidates, named KOI 55.01 and KOI 55.02, from the most probable configuration for this system (Table 1). These parameters suggest that the two planets are smaller than Earth, which would make them the smallest planets so far found around a star still undergoing nuclear fusion. Considering their sizes and orbital parameters, their nature is most probably telluric. This conclusion may be further supported by a possible brightness excess (but only a 1.9σ detection at this stage) at phase zero in the phase curve of KOI-55.01 (Fig. 1). This excess, if real, could be associated with an opposition surge, a phenomenon observed with solid bodies having a tenuous (or no) atmosphere²⁰. The two planet candidates are also remarkable for the very high temperatures expected at their surfaces, indicating that they must be undergoing significant evaporation. However, these atypical properties, compared with those of known planets orbiting main-sequence stars, should not be surprising, considering the nature of the parent star. Planets in extreme environments have been reported previously²¹.

Our two planets were most probably swallowed by their parent star when it became a red giant, a stage that ended only ~ 18 million years ago¹⁵. They were probably orbiting further away and may have been dragged deep into the red-giant envelope to their current positions²². The fact that they are of equivalent size with orbits close to a 3:2 resonance may be an important factor explaining their survival. Indeed, preliminary *N*-body calculations (not shown) suggest that this system is dynamically stable, as long as the planet masses remain lower than several Earth masses; the system may also be stable when a third small planet is present between KOI 55.01 and KOI 55.02. This is consistent with our finding that small bodies should be involved.

This discovery suggests that planets may influence stellar evolution⁹. The increased envelope mass ejection required to form sdB stars from red giants is expected to occur in close interacting binaries, where the companion can transfer some of its orbital momentum to the expanded red-giant envelope, speeding up its rotation and thus triggering its enhanced dissipation²³. However, this does not explain how about half

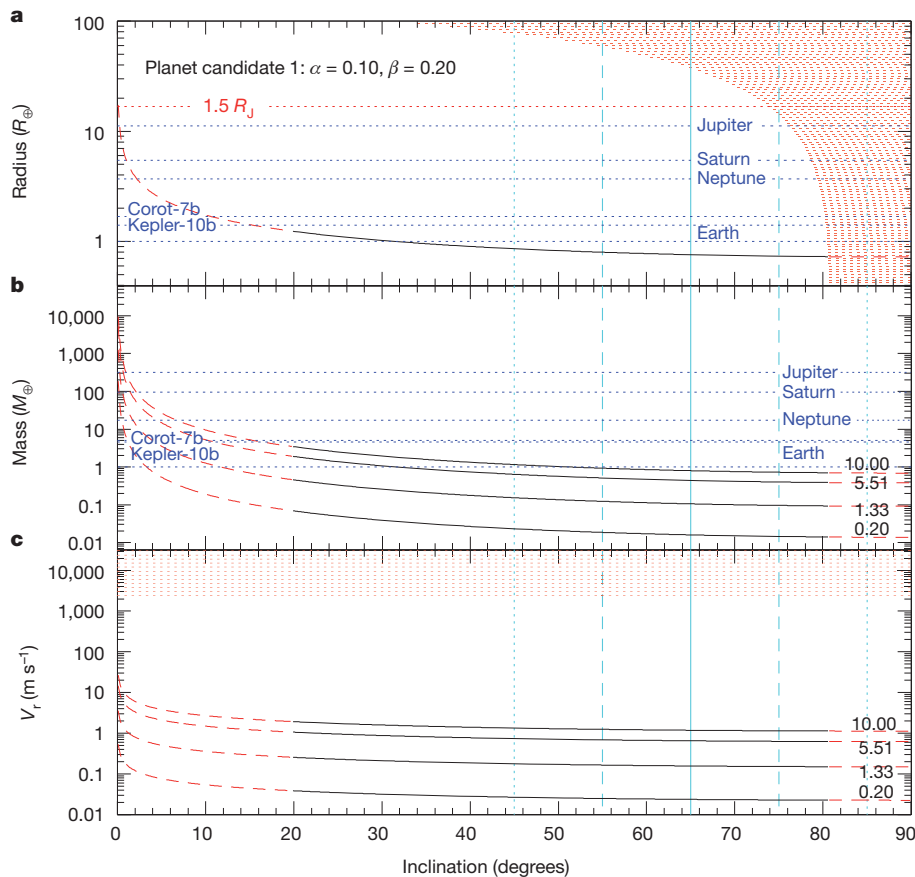


Figure 2 | The calculated properties of the first planet candidate (KOI 55.01) depend on the inclination of the system. Shown are the inclination dependence of the estimated radius (a; in Earth radii, R_{\oplus}), the estimated mass (b; in Earth masses, M_{\oplus}) and the host star projected radial velocity modulation (c; V_r , in m s^{-1}) as functions of the inclination. We assume representative values for the Bond albedo ($\alpha = 0.10$) and the average temperature contrast between the night and day sides ($\beta = 0.2$; that is, approximately the contrast observed on Mercury). The β -parameter strongly depends on the efficiency of heat redistribution on the planet's surface and, consequently, on the presence or absence of an atmosphere. A hot Jupiter with efficient heat redistribution has indeed been found: HD189333b, which, unlike Mercury, has a measured β of 0.8 (ref. 16). However, the precise value of β has a limited impact, as long as $\beta < 0.8$ (Supplementary Information section D). The red filled regions (and dashed portions of the curves) shows excluded domains due to the absence of eclipses (at large inclinations), due to our current upper limit on projected radial velocity modulations (2.4 km s^{-1} at 2σ ; Supplementary Information section C5), and due to asteroseismic indications rejecting an inclination angle

of the sdB stars that, like KIC 05807616, appear to be single or in well detached binaries²⁴, could be produced. In the present case, the planets may have contributed to the dissipation of the envelope to the point of making KIC 05807616 become an extreme horizontal branch star instead of a normal horizontal branch star with a thicker residual envelope²⁵. A plausible scenario would be that these bodies were originally giant planets immersed in the red-giant envelope and massive enough to survive engulfment and trigger the enhanced mass loss necessary for the formation of a hot B subdwarf star²⁶. During that episode, the planets may have been stripped down, losing their gaseous layers and being left only with their inner rocky/iron cores, which would be exposed. The small detected bodies would then be these cores, also named Chthonian planets²⁷, which are dense enough to lie beyond their Roche limit (see Table 1), thus avoiding disruption by the strong tides generated by the parent star.

Alternative scenarios may also be considered. Another way to form single sdB stars is through the merger of two helium white dwarfs²³, and planet formation following this event may be possible²⁸. We could

lower than $\sim 20^\circ$, instead suggesting that $i \approx 65^\circ \pm 10^\circ$. This last constraint is indicated by the light-blue vertical lines that show the expected angle (65° , solid line) with its approximate 1σ (dashed lines) and 2σ (dotted lines) ranges (Supplementary Information section A4). The radii and masses of Jupiter, Saturn, Neptune, Corot-7b²⁹ and Kepler-10b³⁰, are indicated for comparison purposes (dark-blue dotted horizontal lines). We also indicate the limit of $1.5R_J$ (R_J , Jupiter's radius; the red dotted horizontal line in a). Four different curves are shown for the mass and projected radial velocity estimations which sample the typical range of density, ρ_p , where extrasolar planets are usually found. The values $\rho_p = 5.51 \text{ g cm}^{-3}$ and $\rho_p = 1.33 \text{ g cm}^{-3}$ correspond to the mean density of Earth and Jupiter, respectively. With all these considerations, the most likely solution for the planetary parameters clearly points towards planets slightly smaller than our Earth, and therefore most probably of a solid nature (hence with a high mean density; see Table 1). A dynamical confirmation through a measurement of the projected radial velocity of the host star remains difficult at this stage for this relatively faint ($V = 14.87 \text{ mag}$) star, but may be possible with the next generation instrumentation.

speculate that the collapse of the extended envelope resulting from this merger could produce a circumstellar disk, where second generation planets may form. However, it seems unlikely that new, sufficiently dense, planets could have formed within a rather short period of time (less than $\sim 18 \text{ Myr}$) in an environment that close to this hot star.

Received 6 July; accepted 13 October 2011.

1. Udry, S. & Santos, N. C. Statistical properties of exoplanets. *Annu. Rev. Astron. Astrophys.* **45**, 397–439 (2007).
2. Setiawan, J. *et al.* A giant planet around a metal-poor star of extragalactic origin. *Science* **330**, 1642–1644 (2010).
3. Silvotti, R. *et al.* A giant planet orbiting the 'extreme horizontal branch' star V391 Pegasi. *Nature* **449**, 189–191 (2007).
4. Lee, J. W. *et al.* The sdB+M eclipsing system HW Virginis and its circumbinary planets. *Astron. J.* **137**, 3181–3190 (2009).
5. Maxted, P. F. L., Napiwotzki, R., Dobbie, P. D. & Burleigh, M. R. Survival of a brown dwarf after engulfment by a red giant star. *Nature* **442**, 543–545 (2006).
6. Geier, S. *et al.* Binaries discovered by the MUCHFUSS project: SDSS J08205+0008—an eclipsing subdwarf B binary with a brown dwarf companion. *Astrophys. J.* **731**, L22–L26 (2011).

7. D'Cruz, N. L., Dorman, B., Rood, R. T. & O'Connell, R. W. The origin of extreme horizontal branch stars. *Astrophys. J.* **466**, 359–371 (1996).
8. Heber, U. Hot subdwarf stars. *Annu. Rev. Astron. Astrophys.* **47**, 211–251 (2009).
9. Soker, N. Can planets influence the horizontal branch morphology? *Astron. J.* **116**, 1308–1313 (1998).
10. Gilliland, R. L. *et al.* Kepler asteroseismology program: introduction and first results. *Publ. Astron. Soc. Pacif.* **122**, 131–143 (2010).
11. Østensen, R. H. *et al.* First Kepler results on compact pulsators – I. Survey target selection and the first pulsators. *Mon. Not. R. Astron. Soc.* **409**, 1470–1486 (2010).
12. Reed, M. D. *et al.* First Kepler results on compact pulsators – III. Subdwarf B stars with V1093 Her and hybrid (DW Lyn) type pulsations. *Mon. Not. R. Astron. Soc.* **409**, 1496–1508 (2010).
13. Green, E. M. *et al.* Discovery of a new class of pulsating stars: gravity-mode pulsators among subdwarf B stars. *Astrophys. J.* **583**, L31–L34 (2003).
14. Charpinet, S. *et al.* Progress in sounding the interior of pulsating hot subdwarf stars. *AIP Conf. Proc.* **1170**, 585–596 (2009).
15. Van Grootel, V. *et al.* Early asteroseismic results from Kepler: structural and core parameters of the hot B subdwarf KPD 1943+4058 as inferred from g-mode oscillations. *Astrophys. J.* **718**, L97–L101 (2010).
16. Knutson, H. A. *et al.* A map of the day-night contrast of the extrasolar planet HD 189733b. *Nature* **447**, 183–186 (2007).
17. Snellen, I. A. G., Mooij, E. J. W. & Albrecht, S. The changing phases of extrasolar planet CoRoT-1b. *Nature* **459**, 543–545 (2009).
18. Leigh, C. *et al.* A new upper limit on the reflected starlight from τ Bootis b. *Mon. Not. R. Astron. Soc.* **344**, 1271–1282 (2003).
19. Rowe, J. F. *et al.* The very low albedo of an extrasolar planet: MOST space-based photometry of HD209458. *Astrophys. J.* **689**, 1345–1353 (2008).
20. Mallama, A. Characterization of terrestrial exoplanets based on the phase curves and albedos of Mercury, Venus and Mars. *Icarus* **204**, 11–14 (2009).
21. Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature* **355**, 145–147 (1992).
22. Villaver, E. & Livio, M. The orbital evolution of gas giant planets around giant stars. *Astrophys. J.* **705**, L81–L85 (2009).
23. Han, Z. *et al.* The origin of subdwarf B stars – I. The formation channels. *Mon. Not. R. Astron. Soc.* **336**, 449–466 (2002).
24. Maxted, P. F. L. *et al.* The binary fraction of extreme horizontal branch stars. *Mon. Not. R. Astron. Soc.* **326**, 1391–1402 (2001).
25. Bear, E. & Soker, N. Connecting planets around horizontal branch stars with known exoplanets. *Mon. Not. R. Astron. Soc.* **411**, 1792–1802 (2011).
26. Nordhaus, J. *et al.* Tides and tidal engulfment in post-main-sequence binaries: period gaps for planets and brown dwarfs around white dwarfs. *Mon. Not. R. Astron. Soc.* **408**, 631–641 (2010).
27. Hébrard, G., Lecavelier Des Étangs, A., Vidal-Madjar, A., Désert, J.-M. & Ferlet, R. Evaporation rate of hot Jupiters and formation of chthonian planets. *ASP Conf. Proc.* **321**, 203–204 (2004).
28. Silvotti, R. The subdwarf B + giant planet system V391 Peg: Different scenarios for its previous evolution. *ASP Conf. Ser.* **392**, 215–219 (2008).
29. Léger, A. *et al.* Transiting exoplanets from the CoRoT space mission. VIII. CoRoT-7b: the first super-Earth with measured radius. *Astron. Astrophys.* **506**, 287–302 (2009).
30. Batalha, N. M. *et al.* Kepler's first rocky planet: Kepler-10b. *Astrophys. J.* **729**, 27–47 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.C. thanks the Programme National de Physique Stellaire (PNPS, CNRS/INSU, France) for support. G.F. acknowledges the support of the NSERC of Canada and the contribution of the Canada Research Chair Program. MMT spectra for KIC 05807616 were obtained at the MMT Observatory, a joint facility of the University of Arizona and the Smithsonian Institution. A.S.B. acknowledges funding from the Polish Ministry of Science and Higher Education. S.C. thanks R. Gilliland for help in managing Kepler's procedures. We acknowledge the Kepler team and all who have contributed to making this mission possible. Funding for the Kepler mission is provided by NASA's Science Mission Directorate.

Author Contributions S.C. wrote the manuscript, and analysed and interpreted the data from which the presence of planetary bodies was inferred. S.C., G.F., P.B. and V.V.G. derived and checked the calculations to estimate the planets' properties. G.F., P.B. and S.K.R. computed the cut-off frequencies and the theoretical mode visibilities based on a model of the star from V.V.G. E.M.G. obtained and analysed radial velocity measurements. S.D.K., A.S.B., R.H.Ø, R.S. and J.H.T. from the KASC WG11 group independently checked the detection of the orbital signals in the data. All authors discussed the results and contributed to their interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C. (stephane.charpinet@ast.obs-mip.fr).

Orbital excitation blockade and algorithmic cooling in quantum gases

Waseem S. Bakr¹, Philipp M. Preiss¹, M. Eric Tai¹, Ruichao Ma¹, Jonathan Simon¹ & Markus Greiner¹

Interaction blockade occurs when strong interactions in a confined, few-body system prevent a particle from occupying an otherwise accessible quantum state. Blockade phenomena reveal the underlying granular nature of quantum systems and allow for the detection and manipulation of the constituent particles, be they electrons¹, spins², atoms^{3–5} or photons⁶. Applications include single-electron transistors based on electronic Coulomb blockade⁷ and quantum logic gates in Rydberg atoms^{8,9}. Here we report a form of interaction blockade that occurs when transferring ultracold atoms between orbitals in an optical lattice. We call this orbital excitation blockade (OEB). In this system, atoms at the same lattice site undergo coherent collisions described by a contact interaction whose strength depends strongly on the orbital wavefunctions of the atoms. We induce coherent orbital excitations by modulating the lattice depth, and observe staircase-like excitation behaviour as we cross the interaction-split resonances by tuning the modulation frequency. As an application of OEB, we demonstrate algorithmic cooling^{10,11} of quantum gases: a sequence of reversible OEB-based quantum operations isolates the entropy in one part of the system and then an irreversible step removes the entropy from the gas. This technique may make it possible to cool quantum gases to have the ultralow entropies required for quantum simulation^{12,13} of strongly correlated electron systems. In addition, the close analogy between OEB and dipole blockade in Rydberg atoms provides a plan for the implementation of two-quantum-bit gates¹⁴ in a quantum computing architecture with natural scalability.

An ultracold gas of bosonic atoms in the ground band of an optical lattice is described by the Bose–Hubbard model¹⁵, in which atoms can tunnel between neighbouring sites and interact through an onsite repulsive contact interaction. In a deep lattice where the interactions dominate, the ground state of the system is a Mott insulator with a fixed atom number per site that is locally constant over a region of the insulator¹⁶. The energy per site in the absence of tunnelling is $U_{\text{gg}}n(n-1)/2$, where U_{gg} is the interaction energy for two atoms in the ground (lowest-energy) lattice orbital state and n is the number of atoms at the site. The Mott state shows a transport blockade phenomenon in which the presence of an atom at a site energetically prevents tunnelling of a neighbouring atom to that site even in the presence of a small bias between the sites. The transport is blocked unless the bias makes up for the interaction cost, making it possible, for example, to count atoms tunnelling across double wells in a superlattice³. In this work, we explore an excitation blockade phenomenon that does not involve transport in the lattice. The excitation transfers localized atoms between different orbitals at the same site through modulation of the lattice depth at a frequency close to a vibrational resonance. Physics in higher-energy orbitals of the optical lattice has been the focus of much recent experimental work, including the study of dynamics in higher orbitals¹⁷, multi-orbital corrections to the interaction energy¹⁸ and unconventional forms of superfluidity involving higher-energy orbitals^{19,20}.

The OEB mechanism can be understood in the simplest situation for two atoms in a single site of a deep, three-dimensional lattice in

which the vibrational frequencies in all three directions are taken to be different to avoid degeneracies. The lattice depth along the z direction is modulated weakly, which in the presence of anharmonicity of the lattice potential drives atoms between the ground orbital and a single, specific excited z orbital, subject to a selection rule that allows coupling only to orbitals of the same symmetry. For a single atom, excitation to the m th orbital requires modulation at a frequency $\omega_{z,0 \rightarrow m}$, which, ignoring the anharmonicity of the onsite potential, is approximately $m\omega_{z,0 \rightarrow 1}$. If there is more than one atom at a site, the interaction introduces an orbital-dependent shift of the energy levels (Fig. 1). In general, the interaction shifts U_{gg} , U_{ge} and U_{ee} (where the subscripts ‘g’ and ‘e’ denote atoms in the ground and excited orbitals, respectively) are all different and the differences are a significant fraction of U_{gg} . If the coupling strength due to the modulation is small relative to these differences and the modulation frequency is tuned to $\omega_{z,0 \rightarrow m} + (U_{\text{ge}} - U_{\text{gg}})/\hbar$, where \hbar denotes Planck’s constant (\hbar divided by 2π), only a single atom is transferred to the higher orbital and the transfer of a second atom is off resonance. In this sense, the first excitation blocks the creation of a second excitation.

The experimental system has been described in previous work²¹. A two-dimensional Bose–Einstein condensate of ⁸⁷Rb atoms resides in a single plane of a one-dimensional optical lattice, henceforth referred to as the axial lattice, with a vibrational frequency of $\omega_{z,0 \rightarrow 1} = 2\pi \times 5.90(2)$ kHz, where the number in parentheses denotes s.e.m. The z axis is perpendicular to the plane and points in the direction of gravity. In addition, we introduce a lattice in the plane with a spacing of $a = 680$ nm and a depth of $45E_r$ (trap frequency, 17 kHz), where $E_r = \hbar^2/8m_{\text{Rb}}a^2$ is the

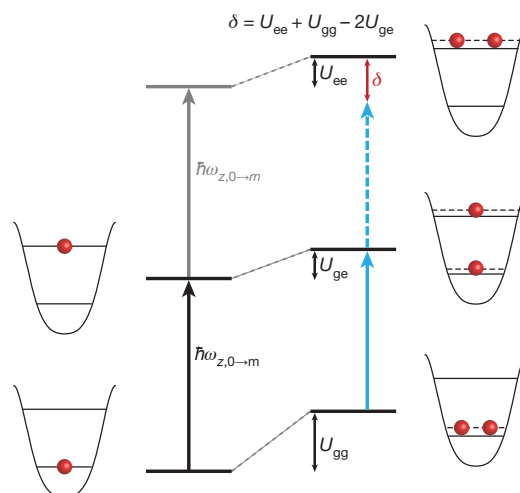


Figure 1 | Orbital excitation blockade mechanism in an optical lattice. A single atom at a site is excited to a higher-energy orbital by resonantly modulating the lattice depth. For two atoms at the same site, interactions lead to an orbital-dependent energy shift. Modulation at the appropriate frequency excites one of the atoms to the higher orbital but is off resonance for exciting the second, differing from resonance by a blockade energy δ .

¹Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

recoil energy of the effective lattice wavelength with m_{Rb} the mass of ^{87}Rb . The resulting Mott insulator is at the focus of a high-resolution microscope capable of detecting atoms at individual lattice sites by fluorescence imaging. Light-assisted collisions that eject pairs of atoms from each lattice site at the start of the imaging process reduce the occupation of a site to its odd-even parity²¹.

We start by demonstrating coherent driving of atoms in a Mott insulator between two orbitals. In the presence of a harmonic trap, the atoms in a two-dimensional Mott insulator are arranged in concentric rings of fixed atom number per site, known as shells, with the greatest occupation at the centre²¹. We prepare a Mott insulator with two shells and modulate the axial lattice depth by $\pm 1.1(1)\%$ at a frequency chosen to transfer atoms from the ground orbital to the second excited orbital. A modulation frequency corresponding to exactly $\omega_{z,0\rightarrow 2}$ is resonant for atoms in the outer shell with one atom per site ($n = 1$). Excitation to the fourth excited orbital is suppressed because of an energy shift of $h \times 1,200(80)$ Hz due to the anharmonicity of the onsite potential. Rabi oscillations between the ground orbital, $|g\rangle$, and an excited orbital, $|e\rangle$, are detected by lowering the axial lattice depth at the end of the modulation such that the excited orbital state becomes unbound and any population in it escapes along the z axis owing to gravity. The Rabi oscillations in that shell (Fig. 2b) have a frequency $\Omega = 2\pi \times 23.3(2)$ Hz.

The OEB is demonstrated in the inner $n = 2$ shell by modulating at a frequency of $\omega_{z,0\rightarrow 2} + (U_{\text{ge}} - U_{\text{gg}})/h$. For our parameters, U_{gg} , U_{ge} and U_{ee} are $h \times 480(30)$, $h \times 360(20)$ and $h \times 310(20)$ Hz, respectively. The Rabi oscillations between $|g, g\rangle$ (the state in which both atoms are in the ground orbital) and $(|e, g\rangle + |g, e\rangle)/\sqrt{2}$ (Fig. 2c) are detected as an oscillation in the parity between even and odd after ejecting the atom in the excited orbital. For the same modulation amplitude as

before, the oscillations are expected to occur $\sqrt{2}$ times faster than the resonant oscillation in the $n = 1$ shell, owing to Bose enhancement⁵. We indeed observe a frequency ratio of 1.42(1) between the oscillation frequencies. A full frequency spectrum in the two shells is shown in Fig. 2a and the frequency separation of the two resonances, 160(10) Hz, matches the theoretically expected value of 165(35) Hz when the effect of virtual orbital-changing collisions is included (Methods).

We next use OEB to demonstrate a new means of cooling quantum gases. Evaporative cooling is the standard technique for cooling atomic gases to nanokelvin temperatures. However, present interest in studying the physics of strongly correlated materials, such as high-transition-temperature copper oxides, using ultracold gases^{12,13} has spurred research into developing new cooling techniques that can reach the requisite picokelvin regime^{22,23}. The field of quantum information offers an alternative cooling method, whereby a sequence of unitary quantum gates purifies a subset of the quantum bits (qubits) in a system by moving entropy and isolating it in another part of the system¹⁰. One realization of such a cooling scheme, heat-bath algorithmic cooling, has been successfully demonstrated with solid-state NMR qubits¹¹. We introduce an analogous technique for quantum gases whereby the unitary operations are achieved using OEB, building on previous theoretical proposals in this direction^{24–27}.

A bosonic quantum gas at a finite temperature, T , adiabatically loaded into the ground band of an optical lattice stores its entropy in the form of atom number fluctuations in the zero-tunnelling limit. Within the local density approximation, a lattice site with a local chemical potential μ is described by a density matrix $\hat{\rho} = \sum_n p_n |n\rangle\langle n|$, where p_n , the probability of having n atoms at the site, is given by $e^{-\beta(U_{\text{gg}}n(n-1)/2 - \mu n)} / Z$. Here $\beta = 1/k_{\text{B}}T$, where k_{B} is Boltzmann's constant, and Z is the grand canonical partition function. Cooling to zero temperature is achieved by changing the atom number distribution on each site to obtain $p_n = \delta_{n, \lceil \mu/U_{\text{gg}} \rceil}$, where δ denotes the Kronecker delta and $\lceil x \rceil$ denotes the least integer greater than x .

The crucial ingredient for algorithmic cooling is a unitary operation that realizes the transformation $|n, m\rangle \rightarrow |n-1, m+1\rangle$ for each n separately, where $|n, m\rangle$ denotes a Fock state with n atoms in the ground orbital and m atoms in an excited orbital. Resonant lattice modulation in the presence of OEB results in a rotation gate $\hat{R}_{nm}(t) = \exp[i(\Omega_{nm}t|n-1, m+1\rangle\langle n, m| + \text{h.c.})]$, where Ω_{nm} is the transition's Rabi frequency, 'h.c.' denotes the Hermitian conjugate terms and the required transformation is achieved for a modulation time $t = \pi/2\Omega_{nm}$. Entropy is transferred from the ground band to the excited band by performing a sequence of π -rotation gates $\hat{R}_{N-s,s}$ for $s = 0$ to $N-1$, with N chosen large enough that $p_N \approx 0$ in the initial state. At the end of this sequence, most of the entropy of the gas has been transferred to the excited band and can be removed from the system by ejecting the atoms in that band. The local chemical potential, μ , is re-adjusted to recover a situation closer to thermal equilibrium by reducing the harmonic confinement frequency to a new value, ω_{low} , such that $\mu < U_{\text{gg}}$ throughout the gas. At this point, residual entropy is stored in the resulting $n = 1$ Mott insulator in the form of holes that are preferentially located near the edge of the sample. The gas is made to rethermalize by lowering the lattice depth to allow tunnelling, and the entropy of the final, thermalized, state should be significantly lower than that of the initial state.

We start by experimentally demonstrating the algorithm from a state with known atom number, namely a four-shell Mott insulator, and reducing the site occupation everywhere in the insulator to one atom per site. To increase the blockade energy for this set of experiments, we transfer atoms to the fourth axial orbital rather than the second. We also replace the rotation gates used above with Landau–Zener transitions to improve the fidelity of the algorithm. We linearly sweep the modulation frequency from 20.90 to 21.65 kHz in 250 ms. This chirp realizes a sequence of quantum operations that transfer atoms to the

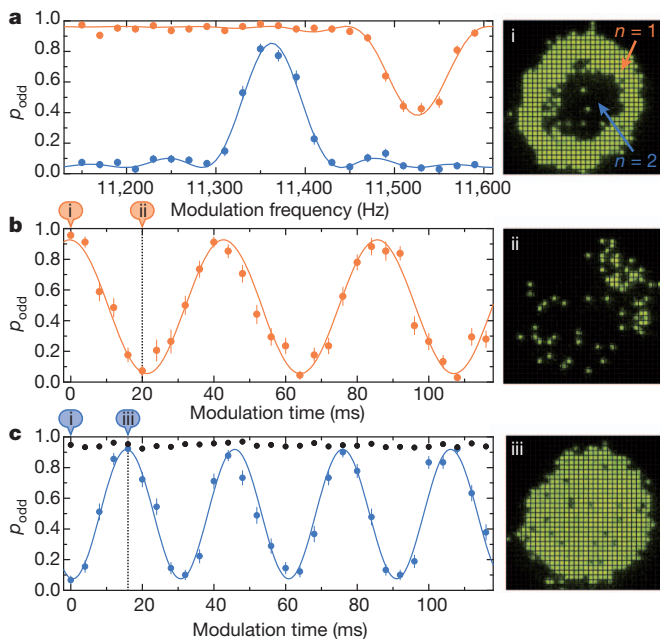


Figure 2 | Time-, frequency- and site-resolved coherent transfer of atoms between orbitals in a Mott insulator. **a**, Excitations transferring a single atom in the $n = 1$ (orange) or $n = 2$ shell (blue) from the ground orbital to the second excited orbital are spectroscopically resolved in a two-shell Mott insulator. P_{odd} , probability of odd occupancy at a site. **b**, **c**, Rabi oscillations between the two orbitals are observed by driving at the resonant frequencies for atoms in the $n = 1$ (**b**) and $n = 2$ (**c**) shells of a Mott insulator. Bose enhancement leads to faster oscillations in the $n = 2$ shell. When the atom number is reduced to one atom per site in the region previously containing two atoms, the interaction shift suppresses oscillations (black). Error bars, 1 s.e.m. The plots on the right are site-resolved snapshots of the Mott insulator at different points in the Rabi cycles (i–iii).

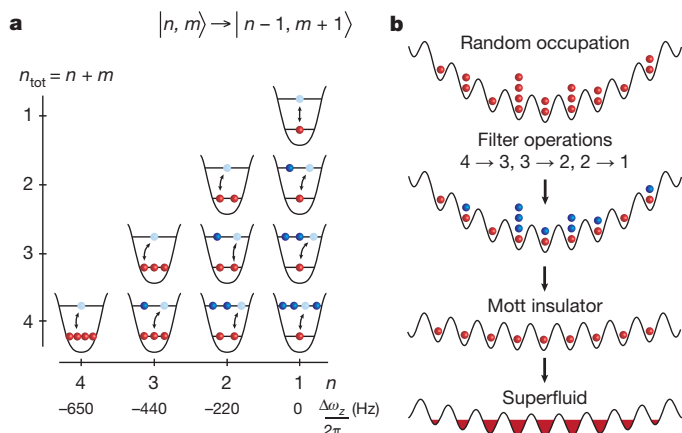


Figure 3 | Algorithmic cooling in an optical lattice. **a**, Landau-Zener chirp for transferring entropy from the ground band to the fourth excited band. The lattice modulation frequency is swept across the transition resonances from left to right. The values of the interaction shift, $\Delta\omega_z$, between the excitation frequency for one of n atoms transferred to the fourth orbital and the excitation frequency for a single atom at a site are shown for different orbital occupancies. Excitation processes in the same column happen at the same frequency, to within 30 Hz (Supplementary Table 1). **b**, A state with random occupation in a deep lattice is far from the Mott-insulating ground state. Sequential filtering operations and subsequent reduction of the confinement prepare the ground state, which can be adiabatically converted to a thermalized superfluid in a shallow lattice. Red and blue spheres denote atoms in the ground and excited bands, respectively.

excited orbital one at a time, until only one atom remains in the ground orbital in all shells (Fig. 3a). We probe the ground-orbital occupancy at different points in the frequency chirp by ejecting atoms from the higher orbital as before and then performing the parity imaging. The parity of the different shells during the chirp is shown in Fig. 4a, and an analogue of the typical Coulomb blockade staircase is seen in the data. Shell-sensitive manipulation of a Mott insulator was achieved in previous experiments using a microwave transition between hyperfine states, but the lack of a strong blockade allowed transfer of only a small fraction of the population to the target state²⁸.

Next we demonstrate cooling by performing the algorithm on a state that is far from the many-body ground state. To prepare such a state, we non-adiabatically load a condensate into a deep lattice, projecting the wavefunction onto a state with Poissonian site occupancy that rapidly loses coherence between sites. Using the same operation sequence as before, we progressively reduce the randomness of the

ground-band occupancy, preparing a single-occupancy Mott insulator (Fig. 3b). We increase the odd-parity occupancy from 0.45(1) to 0.76(2) (Fig. 4b), demonstrating significant atom number squeezing limited by the conversion efficiency of the Landau-Zener transitions. To complete the algorithm, we re-adjust the harmonic confinement to obtain a state close to the many-body ground state. We verify the ground-state character by ramping back adiabatically to a $5E_r$ lattice in 100 ms and releasing the atoms from the lattice. Without cooling, we obtain a featureless cloud (Fig. 4c, top), indicating an absence of the coherence expected in the superfluid ground state. With cooling, the Mott insulator is adiabatically converted to a superfluid, giving rise to matter-wave interference peaks (Fig. 4c, bottom).

We now discuss the limits on the entropies that can be achieved with algorithmic cooling. The efficiency of conversion to a single-occupancy Mott insulator is technically limited in our system by heating due to spontaneous emission during the sweep and, to a lesser extent, by the efficiency of the Landau-Zener sweep for clouds (samples) with large average occupancies (Methods). Although we have demonstrated cooling of hot clouds, the single-occupancy probability we have achieved using algorithmic cooling in a two-shell Mott insulator is 0.94(1). This is comparable to what was previously achieved with evaporative cooling, corresponding to an average entropy of $0.27k_B$ per particle²¹. Nevertheless, lattice heating can be made negligible by using a further-detuned lattice (for example with a wavelength of 1,064 nm), and shaped pulses can improve the Landau-Zener transfer efficiency²⁵. More fundamentally, the single-shot cooling algorithm we have implemented is limited by initial holes in the Mott insulator that cannot be corrected. However, repeated iterations of the algorithm can circumvent this problem and bring the cloud to zero entropy rapidly²⁵. The cycle alternates between using OEB to produce a reduced-entropy, $n = 1$ insulator in weak harmonic confinement (demonstrated above) and adiabatically increasing the confinement in the presence of tunnelling to move hot particles to the centre of the cloud, where they can be removed again. Alternatively, the outer edge of the cloud containing the holes can be removed using the high resolution available in our system²⁹.

We have observed a new blockade phenomenon in optical lattices when exciting atoms to higher orbitals, analogous to dipole excitation blockade in Rydberg atoms. The blockade permits deterministic manipulation of atom number in an optical lattice. We have used it to convert a multi-shell Mott insulator into a singly occupied insulator with over 1,000 sites, the largest quantum register achieved so far in an addressable system. The same technique allows initialization of registers in longer-wavelength lattices where a Mott insulator cannot

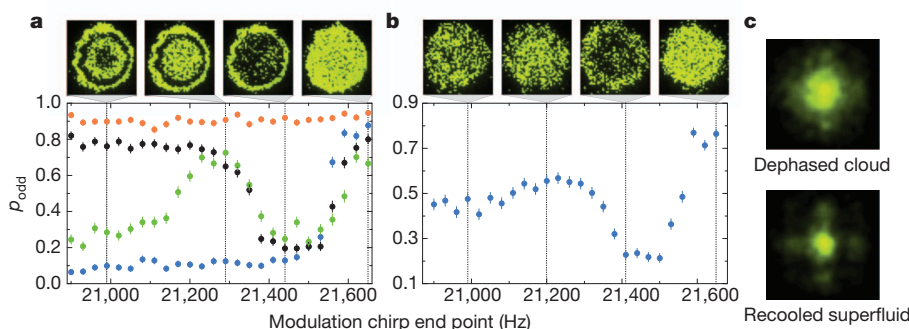


Figure 4 | Experimental realization of algorithmic cooling. **a**, By chirping the modulation towards higher frequencies, atoms in a four-shell Mott insulator are sequentially excited one at a time to the fourth orbital. The population in the higher-energy orbital is subsequently ejected at the end of the chirp. The average parities in the $n = 1$ (orange), $n = 2$ (blue), $n = 3$ (black) and $n = 4$ (green) shells are shown at different points in the chirp, together with single-shot images, illustrating the conversion to a three-, a two- and, finally, a one-shell insulator. **b**, The same frequency chirp algorithmically cools a state

with random occupancy into an $n = 1$ Mott insulator, observed as an enhancement in odd-parity occupancy. Error bars, 1 s.e.m. **c**, Top: after the lattice depth is adiabatically lowered, an incoherent cloud does not show an interference pattern in a expansion with a 5-ms time of flight. Bottom: cooling converts the incoherent cloud to a Mott insulator in the deep lattice. After the lattice depth is adiabatically lowered, a superfluid forms and an interference pattern is obtained in the expansion images.

be prepared^{30,31}. OEB also opens a route to implementing quantum gates in optical lattices. Single-site addressing²⁹, possible with our microscope, can perform rotations of individual orbital-encoded qubits rather than the global rotations demonstrated in this work. Controlled-NOT gates can be implemented by conditionally moving the control qubit to the target qubit site and performing an interaction-sensitive rotation of the target qubit¹⁴. Finally, the algorithmic cooling technique we have developed could potentially achieve the ultralow entropies required for quantum simulation^{12,13} and computation in optical lattices, and establishes a connection with quantum information through which may come novel ideas for cooling quantum gases.

METHODS SUMMARY

Our experiments start with a single-layer, two-dimensional quantum gas of ⁸⁷Rb in a square lattice with a spacing of 680 nm. Confinement in the plane is achieved by putting the atoms in a single well of an axial lattice with a trap frequency of $2\pi \times 5.90$ kHz. The atom plane is at the focus of a high-resolution imaging system with a numerical aperture of 0.8, capable of resolving atoms at individual sites. To suppress any tunnelling when studying the OEB process, the in-plane lattice depth is ramped to $45E_r$. To drive transitions between orbitals, the axial lattice depth is modulated at the appropriate frequency, taking into account interaction shifts. Atoms in higher orbitals are removed by decreasing the depth of the axial lattice, allowing the atoms to Zener tunnel away in a few milliseconds owing to the gravitational potential in that direction. The site occupations of the remaining ground-orbital atoms are then detected using fluorescence imaging. The depths of the in-plane and axial lattices are rapidly increased to 300 μ K and 3 mK, respectively, to pin the atoms during the imaging process, and red-detuned optical molasses beams serve to illuminate and cool the atoms simultaneously. Light-assisted collisions rapidly eject atoms in pairs, and photons scattered by residual atoms at sites that had an odd occupation number are detected on a charge-coupled device.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 May; accepted 18 October 2011.

- Grabert, H. & Devoret, M. H. (eds). *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* 21–137 (Springer, 1992).
- Ono, K., Austing, D. G., Tokura, Y. & Tarucha, S. Current rectification by Pauli exclusion in a weakly coupled double quantum dot system. *Science* **297**, 1313–1317 (2002).
- Cheinet, P. *et al.* Counting atoms using interaction blockade in an optical superlattice. *Phys. Rev. Lett.* **101**, 090404 (2008).
- Urban, E. *et al.* Observation of Rydberg blockade between two atoms. *Nature Phys.* **5**, 110–114 (2009).
- Gaëtan, A. *et al.* Observation of collective excitation of two individual atoms in the Rydberg blockade regime. *Nature Phys.* **5**, 115–118 (2009).
- Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
- Kastner, M. A. The single-electron transistor. *Rev. Mod. Phys.* **64**, 849–858 (1992).
- Isenhower, L. *et al.* Demonstration of a neutral atom controlled-NOT quantum gate. *Phys. Rev. Lett.* **104**, 010503 (2010).
- Wilk, T. *et al.* Entanglement of two individual neutral atoms using Rydberg blockade. *Phys. Rev. Lett.* **104**, 010502 (2010).
- Boykin, P., Mor, T., Roychowdhury, V., Vatan, F. & Vrijen, R. Algorithmic cooling and scalable NMR quantum computers. *Proc. Natl Acad. Sci. USA* **99**, 3388–3393 (2002).
- Baugh, J., Moussa, O., Ryan, C., Nayak, A. & Laflamme, R. Experimental implementation of heat-bath algorithmic cooling using solid-state nuclear magnetic resonance. *Nature* **438**, 470–473 (2005).
- Lewenstein, M. *et al.* Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond. *Adv. Phys.* **56**, 243–379 (2007).
- Bloch, I., Dalibard, J. & Zwierger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
- Schneider, P. & Saenz, A. Quantum computation with ultracold atoms in a driven optical lattice. Preprint at (<http://arxiv.org/abs/1103.4950>) (2011).
- Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
- Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
- Müller, T., Fölling, S., Widera, A. & Bloch, I. State preparation and dynamics of ultracold atoms in higher lattice orbitals. *Phys. Rev. Lett.* **99**, 200405 (2007).
- Will, S. *et al.* Time-resolved observation of coherent multi-body interactions in quantum phase revivals. *Nature* **465**, 197–201 (2010).
- Wirth, G., Ölschläger, M. & Hemmerich, A. Evidence for orbital superfluidity in the p-band of a bipartite optical square lattice. *Nature Phys.* **7**, 147–153 (2011).
- Soltan-Panahi, P. *et al.* Multi-component quantum gases in spin-dependent hexagonal lattices. *Nature Phys.* **7**, 434–440 (2011).
- Bakr, W. S. *et al.* Probing the superfluid-to-Mott insulator transition at the single-atom level. *Science* **329**, 547–550 (2010).
- McKay, D. C. & DeMarco, B. Cooling in strongly correlated optical lattices: prospects and challenges. *Rep. Prog. Phys.* **74**, 054401 (2011).
- Medley, P., Weld, D. M., Miyake, H., Pritchard, D. E. & Ketterle, W. Spin gradient demagnetization cooling of ultracold atoms. *Phys. Rev. Lett.* **106**, 195301 (2011).
- Rabl, P., Daley, A. J., Fedichev, P. O., Cirac, J. I. & Zoller, P. Defect-suppressed atomic crystals in an optical lattice. *Phys. Rev. Lett.* **91**, 110403 (2003).
- Popp, M., Garcia-Ripoll, J.-J., Vollbrecht, K. G. & Cirac, J. I. Ground-state cooling of atoms in optical lattices. *Phys. Rev. A* **74**, 013622 (2006).
- Nikolopoulos, G. M. & Petrosyan, D. Atom-number filter in an optical lattice. *J. Phys. B* **43**, 131001 (2010).
- Sherson, J. F. & Mølmer, K. Shaking the entropy out of a lattice: atomic filtering by vibrational excitations. Preprint at (<http://arxiv.org/abs/1012.1457>) (2010).
- Campbell, G. K. *et al.* Imaging the Mott insulator shells by using atomic clock shifts. *Science* **313**, 649–652 (2006).
- Weitenberg, C. *et al.* Single-spin addressing in an atomic Mott insulator. *Nature* **471**, 319–324 (2011).
- Nelson, K. D., Li, X. & Weiss, D. Imaging single atoms in a three-dimensional array. *Nature Phys.* **3**, 556–560 (2007).
- Weiss, D. S. *et al.* Another way to approach zero entropy for a finite system of atoms. *Phys. Rev. A* **70**, 040302 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank S. Fölling for discussions. This work was supported by grants from the US Army Research Office with funding from the DARPA OLE program, an AFOSR MURI programme and by grants from the US NSF.

Author Contributions All authors contributed to the construction of the experiment, the collection and analysis of the data, and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.G. (greiner@physics.harvard.edu).

METHODS

State preparation. Our experiments begin with a degenerate, two-dimensional Bose gas of ^{87}Rb atoms prepared in the $|F = 1, m_F = -1\rangle$ state in a single layer of a one-dimensional optical lattice with spacing $1.5\text{ }\mu\text{m}$, in the focal plane of a high-resolution imaging system as described in previous work. The atoms are then loaded into a two-dimensional optical lattice with spacing 680 nm , which is ramped up to a depth of $45E_r$ adiabatically on either a single- or many-body timescale, depending upon the experiment to be performed.

Higher-orbital removal. The orbital blockade is observed through the deterministic removal of atoms from higher-energy orbitals of the $1.5\text{-}\mu\text{m}$ lattice. Removal of atoms from the second excited orbital is achieved by reducing the depth of this lattice to 3.8 kHz from an initial depth of 35.8 kHz . Gravity produces a shift of 3.2 kHz per well, thus inducing second-orbital atoms to Zener tunnel away within a few milliseconds. The Landau–Zener tunnelling rate from the ground orbital is given by $\Gamma_{LZ} = (mga/2\pi\hbar)e^{-g_c/g} \approx 12\text{ Hz}$. Here g is the acceleration due to gravitation and $g_c = a\omega_z^2/4$. This effect leads to a loss of ground-state atoms on the per cent level, but can be made negligible by using excitations to the fourth excited orbital.

Orbital-dependent energy shifts. Owing to its large spacing, the $1.5\text{-}\mu\text{m}$ lattice has a recoil energy of only 250 Hz . Consequently, its depth is ~ 150 recoils, and its low-lying eigenstates are, to a good approximation, the Hermite–Gaussian modes of a harmonic oscillator. The interaction energy between particles in orbitals m and n may thus be written in terms of the ground-orbital interaction energy, U_{00} , as

$$U_{nm} = U_{00}(2 - \delta_{nm}) \frac{\int |\psi_m(x)|^2 |\psi_n(x)|^2 dx}{\int |\psi_0(x)|^4 dx}$$

where $\psi_m(x)$ is the normalized m th harmonic oscillator wavefunction. The total interaction shift for M particles in orbital m and N particles in orbital n is thus

$$\frac{M(M-1)}{2} U_{mm} + \frac{N(N-1)}{2} U_{nn} + MNU_{mn}$$

The interactions also produce off-resonance orbital-changing collisions, with a Rabi frequency

$$\Omega_{mn \leftrightarrow pq} = U_{00} C_{mn \leftrightarrow pq} \frac{\int \psi_m(x) \psi_n(x) \psi_p^*(x) \psi_q^*(x) dx}{\int |\psi_0(x)|^4 dx}$$

where $C_{mn \leftrightarrow pq}$ is a combinatoric factor resulting from Bose enhancement. For an energy defect of $\delta_{mn \leftrightarrow pq} \gg \Omega_{mn \leftrightarrow pq}$, this process produces an energy shift of $\Delta_{mn} = -|\Omega_{mn \leftrightarrow pq}|^2 / \delta_{mn \leftrightarrow pq}$. In our experiment, the dominant orbital-changing collision is $|m=0, n=2\rangle \rightarrow |p=1, q=1\rangle$, with a Rabi frequency $\Omega_{02 \leftrightarrow 11} = 2\pi \times 120\text{ Hz}$. In this case, lattice anharmonicity and interaction shifts produce an energy defect of $\delta_{02 \leftrightarrow 11} = 2\pi \times 330\text{ Hz}$, resulting in an additional overall shift in the energy of the $|0, 2\rangle$ state of $\Delta_{02} = -2\pi \times 45\text{ Hz}$.

Limits on entropies achievable with algorithmic cooling. In our system, the main limitation on the entropies achievable with algorithmic cooling is losses during the Landau–Zener chirp. Spontaneous emission after absorption of photons from the lattice leads to excitation of atoms from the ground orbital at a rate of 0.1 s^{-1} . These atoms are quickly lost owing to tunnelling in the higher-energy orbitals in the case of the in-plane lattice or in the orbital-filtering step in the case of the axial lattice. This leads to holes in an $n=1$ Mott insulator (3% during the 250-ms ramp), setting a lower bound on the reachable entropy after thermalization of $\sim 0.15k_B$ per particle. The lattice light is blue-detuned 25 nm from the atomic resonance, and the heating rate can be made negligible by increasing the detuning (for example by using $1,064\text{-nm}$ lattice light). A more fundamental limit on the single-shot cooling algorithm demonstrated here is caused by initial holes in the Mott insulator that cannot be corrected. For our Mott insulators, the hole fraction is on the order of 0.5% , corresponding to a post-thermalization entropy of $\sim 0.06k_B$ per particle. This limit can be overcome by the iterative algorithm described in the text. For Mott insulators with large initial atom numbers per site, it is also important to take into account the efficiency of the Landau–Zener chirp. For excitation to the second orbital, the measured efficiency is $0.94(1)$.

Tuning charge transport in solution-sheared organic semiconductors using lattice strain

Gaurav Giri¹, Eric Verploegen^{1,2}, Stefan C. B. Mannsfeld², Sule Atahan-Evrenk³, Do Hwan Kim¹, Sang Yoon Lee⁴, Hector A. Becerra⁵, Alán Aspuru-Guzik³, Michael F. Toney² & Zhenan Bao¹

Circuits based on organic semiconductors are being actively explored for flexible, transparent and low-cost electronic applications^{1–5}. But to realize such applications, the charge carrier mobilities of solution-processed organic semiconductors must be improved. For inorganic semiconductors, a general method of increasing charge carrier mobility is to introduce strain within the crystal lattice⁶. Here we describe a solution-processing technique for organic semiconductors in which lattice strain is used to increase charge carrier mobilities by introducing greater electron orbital overlap between the component molecules. For organic semiconductors, the spacing between cofacially stacked, conjugated backbones (the π - π stacking distance) greatly influences electron orbital overlap and therefore mobility⁷. Using our method to incrementally introduce lattice strain, we alter the π - π stacking distance of 6,13-bis(triisopropylsilyl)ethynyl pentacene (TIPS-pentacene) from 3.33 Å to 3.08 Å. We believe that 3.08 Å is the shortest π - π stacking distance that has been achieved in an organic semiconductor crystal lattice (although a π - π distance of 3.04 Å has been achieved through intramolecular bonding^{8–10}). The positive charge carrier (hole) mobility in TIPS-pentacene transistors increased from 0.8 cm² V⁻¹ s⁻¹ for unstrained films to a high mobility of 4.6 cm² V⁻¹ s⁻¹ for a strained film. Using solution processing to modify molecular packing through lattice strain should aid the development of high-performance, low-cost organic semiconducting devices.

Inorganic semiconductor films can exhibit lattice strain due to lattice mismatch by heteroepitaxy, with strained films showing increased charge carrier mobility⁶. Recently, the bandgap of graphene has been modified using uniaxial strain, providing a way of tuning electronic properties¹¹. Similarly, straining the crystal lattice of organic semiconductors could provide a way of changing the electronic properties of organic semiconductors. For organic semiconductors, the π - π stacking distance between cofacially stacked molecules critically affects the charge carrier mobility⁷. Therefore, a strained lattice that results in shorter π - π stacking distances can potentially greatly increase the charge transfer integral, which describes the electronic wavefunction overlap between adjacent molecules⁷. At the same time, the displacement distance along other molecular axes is also important in charge transport⁷.

Changing the chemical structure of the organic semiconductor, or the chemical composition of the dielectric interface, has been shown to affect organic semiconductor molecular packing. These are the most common methods used to design high-performance organic-semiconductor devices^{12–16}. Non-synthetic techniques offer an alternative way of controlling the molecular packing. Some organic semiconductors exhibit temperature-induced solid-state transitions in molecular packing, while others exhibit a thin-film packing structure that is different from the bulk molecular packing motif^{17,18}. The small-molecule organic semiconductor TIPS-pentacene has a 'brick-wall' packing motif with a π - π stacking distance of 3.33 Å as both a thin film and a bulk crystal¹⁹. It is

stable in air, processible through solution-based methods and possesses high hole mobility^{20,21}. Here, we report the formation of strained molecular packing in TIPS-pentacene thin films using our previously reported solution-shearing method²². The π - π stacking distance decreased from 3.33 Å to 3.08 Å for the solution-sheared thin film prepared at a shearing speed of 8 mm s⁻¹. The hole mobility was increased from 0.8 cm² V⁻¹ s⁻¹ for unstrained films prepared at a shearing speed of 0.4 mm s⁻¹ up to as high as 4.6 cm² V⁻¹ s⁻¹ in strained thin films prepared at a speed of 2.8 mm s⁻¹.

During the solution-shearing process, a shearing plate drags the solution across a heated substrate while keeping the bulk of the solution between the plate and the substrate, with only the evaporation front exposed (Fig. 1a and Supplementary Fig. 1)²². Fig. 1b–f shows cross-polarized optical microscope images illustrating the texture of thin films as a function of shearing speed, ranging from 0.4 mm s⁻¹ to 8 mm s⁻¹. For a shearing speed of 0.4 mm s⁻¹ (Fig. 1b and Supplementary Fig. 2), oriented, millimetre-wide TIPS-pentacene domains with lengths of up to a centimetre are observed with the long axis parallel to the shearing direction. Shearing at a speed of 1.6 mm s⁻¹ (Fig. 1c) results in domains that are narrower and shorter in length than those resulting from shearing at 0.4 mm s⁻¹. This trend in TIPS-pentacene morphology continues up to a shearing speed of 2.8 mm s⁻¹ (Fig. 1d).

When the shearing speed is raised to 4 mm s⁻¹, a comet-shaped morphology (known as 'transcrystalline') is observed (Fig. 1e). These features are hundreds of micrometres wide and several millimetres long. Such transcrystalline features have been observed for polymers crystallized in a temperature gradient and for small molecules crystallized in a concentration gradient^{23,24}. A concentration gradient between the bulk solution and the evaporation front also forms during solution shearing²⁵. Increasing shearing speed to 8 mm s⁻¹ results in an isotropic, spherulitic film (Fig. 1f); no preferential orientation relative to the shearing direction is observed for the spherulites.

In addition to altering TIPS-pentacene thin film crystallite orientation (texture) and domain sizes, solution-shearing speed also has an effect on the molecular packing in the thin film. Grazing incidence X-ray diffraction (GIXD) experiments were performed to characterize the texture and molecular packing. As the shearing speed increases, the (101) *d*-spacing decreases incrementally, from 7.70 Å to 7.26 Å, and concurrently the (010) *d*-spacing increases incrementally, from 7.83 Å to 8.13 Å (Fig. 2 and Supplementary Table 1). These *d*-spacing changes occur for the in-plane lattice parameters (*a* and *b*) only, because the vertical layer spacing—the (001) *d*-spacing—does not systematically change with shearing speed (Supplementary Fig. 3). To our knowledge, a non-synthetic methodology to induce incremental thin-film lattice strain has not been reported.

To study lattice strain evolution in detail, high-resolution GIXD measurements were performed to determine the crystal structure (Supplementary Fig. 4). Changes in the positions of GIXD peaks is due to the incrementally changing lattice strain in the TIPS-pentacene

¹Department of Chemical Engineering, Stanford University, Stanford, California 94305, USA. ²Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ³Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Display Device Laboratory, Samsung Advanced Institute of Technology, Mt 14-1, Nongseo-dong, Giheung-gu, Yongin-Si, Kyunggi-Do 449-712, South Korea. ⁵Department of Chemistry, Brigham Young University—Idaho, Rexburg, Idaho 83460, USA.

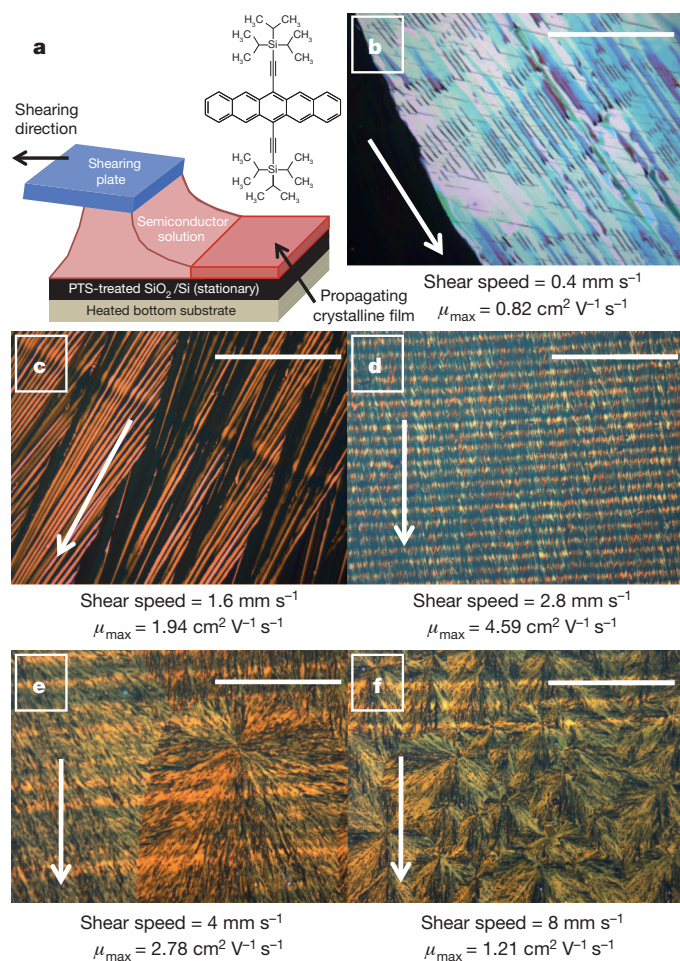


Figure 1 | Solution-shearing schematic and cross-polarized optical microscope images of solution-sheared films. **a**, Schematic diagram of the solution-shearing method. **b–f**, Cross-polarized optical microscope images of solution-sheared TIPS-pentacene thin films, formed with shearing speeds of 0.4 mm s^{-1} , 1.6 mm s^{-1} , 2.8 mm s^{-1} , 4 mm s^{-1} and 8 mm s^{-1} , respectively (scale bars are all $200 \mu\text{m}$). Dark regions of the images are due to crystallites oriented along the polarization direction of the light. In all cases the white arrow represents the shearing direction. The maximum mobility obtained for films with the corresponding thin film crystal texture is given by μ_{max} .

film as a function of solution-shearing speed. The low signal-to-noise ratio at higher shearing speeds is caused by the loss of anisotropy of the crystallite texture. Peak broadening is seen with increasing shearing speed, which is consistent with the much smaller crystallite sizes caused by fast drying at high shearing speeds. It is also possible that the entire film is not uniformly strained, causing a distribution of superimposed peaks. Owing to peak broadening beyond the shearing speed of 4 mm s^{-1} , we cannot exclude the possibility that the peak shift seen from 4 mm s^{-1} to 8 mm s^{-1} is the result of the coexistence of differently strained crystallites. The changes in lattice strain occurring at various shearing speeds are not due to differences in film thickness (Supplementary Fig. 5). We eliminated the possibility of a solvent inclusion complex leading to the observed change in the unit cell dimensions, because heating the strained film up to 160°C did not change the molecular packing, despite this temperature being much higher than the boiling point of the solvent used, toluene. Furthermore, the film sheared at 0.4 mm s^{-1} showed a GIXD pattern identical to the pattern of an evaporated thin film that was not exposed to solvent.

The unit cell geometry of the solution-sheared TIPS-pentacene thin films was obtained from the peak positions measured from the GIXD images using a least-square-error optimization procedure

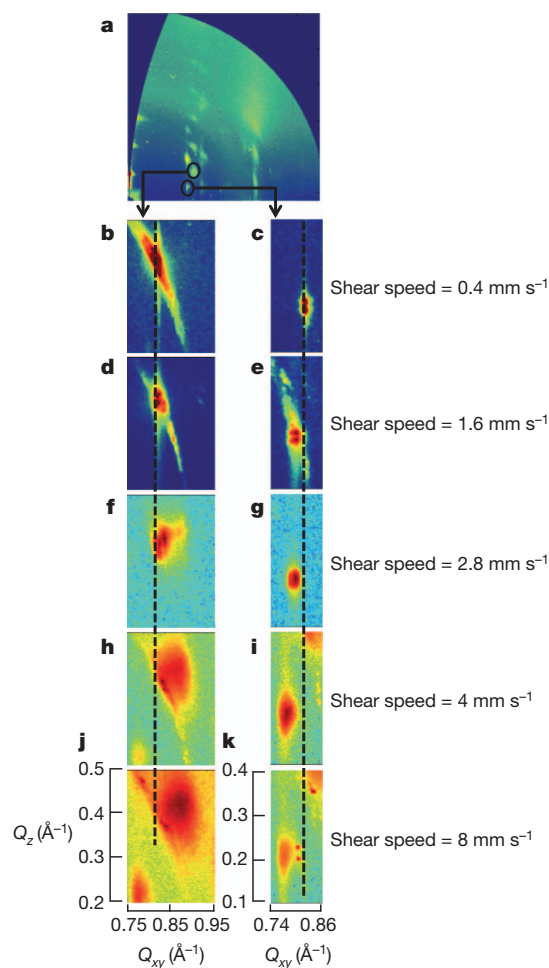


Figure 2 | Change in GIXD pattern of TIPS-pentacene as function of shearing speed. **a**, A representative GIXD pattern of a sheared film of TIPS-pentacene. **b, d, f, h** and **j**, The evolution of the (101) Bragg peak used to calculate the (101) d -spacing of TIPS-pentacene thin films. **c, e, g, i** and **k**, Evolution of the (010) Bragg peak used to calculate the (010) d -spacing of TIPS-pentacene thin films. As shearing speed increases, the (101) Bragg peak shifts to higher q_{xy} (reciprocal spacing) values and the (010) Bragg peak shifts to lower q_{xy} values, signifying unit cell strain in the plane of the substrate surface. The q_{xy} and q_z components are the scattering vectors parallel and perpendicular to the substrate, respectively. A dashed line is placed at the unstrained peak positions to guide the eye.

(Supplementary Table 1)¹⁹. The TIPS-pentacene in-plane unit cell geometry becomes increasingly more oblique with increasing shearing speed, while the in-plane unit cell area remains comparatively constant, which is also consistent with the absence of solvent inclusion in the unit cell (Supplementary Table 1).

To assess the stability of this strained crystal lattice, we exposed the strained films to toluene vapour. This induced a reorganization of the molecular packing without affecting crystallite texture or domain size, indicating that the strained structure is metastable. Upon exposure to toluene vapour a change in GIXD peak position occurred (Supplementary video 1). After an hour of toluene exposure, the lattice spacing of strained films became similar to that of the evaporated thin film, indicating strain relief (Supplementary Fig. 6). On the other hand, the strained films are thermally stable up to 160°C (Supplementary Video 2).

The molecular packing of TIPS-pentacene has been previously determined for evaporated thin films, and was found to be identical to the bulk crystal structure¹⁹. Two distinct molecular pairs, labelled T1 and T2, can be seen (Fig. 3a). Here, the same crystal structure determination technique was applied to films sheared at a speed of

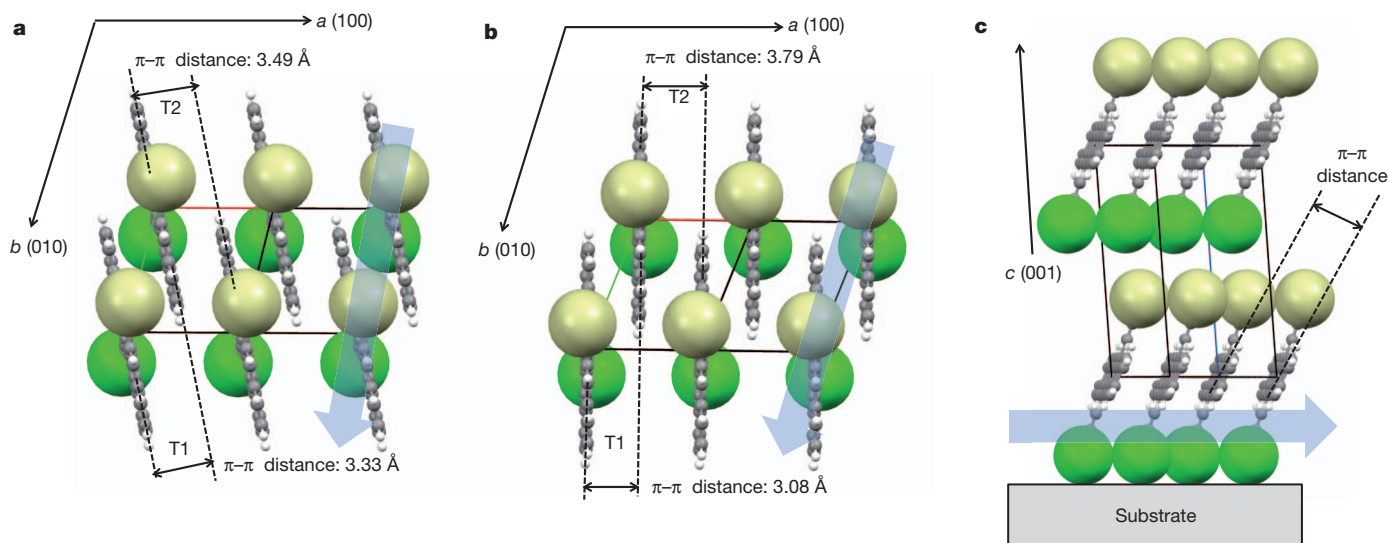


Figure 3 | Molecular packing structure of TIPS-pentacene thin films prepared under different conditions. **a**, Evaporated thin film. **b**, A thin film fabricated from solution shearing at a speed of 8 mm s^{-1} . T1 and T2 denote the unique molecular pairs (Supplementary Table 2). **c**, Solution-sheared thin film

at 8 mm s^{-1} ; the method is described in Supplementary Fig. 7. The crystal structure for the film solution-sheared at 8 mm s^{-1} displayed a more oblique molecular packing motif than did that of the evaporated thin films (Fig. 3b). The π - π stacking distance of the T1 molecular pair (Fig. 3b) decreased from 3.33 \AA in the evaporated thin film to 3.08 \AA in thin films prepared at a shearing speed of 8 mm s^{-1} . This is significant because the π - π stacking distance has an exponential impact on the transfer integral between cofacially arranged π -conjugated organic semiconductor molecules⁷. Density functional theory calculations show that the nearest-neighbour charge transfer integral of the T1 molecular pair increases from 11.7 meV in the evaporated thin film to -36.9 meV in the strained thin film (Supplementary Fig. 8 and Supplementary Table 2). The lattice strain induced in the TIPS-pentacene molecular packing also results in a molecular displacement along the molecular long and short axes. These displacements also affect the charge transfer integral, so the overall charge transfer integral may not necessarily increase with a shorter π - π stacking distance. Density functional theory calculations were performed to check the impact of molecular long and short axis displacements (Supplementary Fig. 9). We also note that the strained T2 molecular pair has a

larger π - π stacking distance of 3.79 \AA , and a low charge transfer integral of 0.429 meV (Fig. 3b and Supplementary Table 2). We hypothesize that charge transport in the strained TIPS-pentacene lattice is one-dimensional and mainly through T1 molecular pairs, primarily along the b axis, which is also the direction of elongated crystallite growth.

We then measured the impact of lattice strain on the electronic performance of TIPS-pentacene thin film transistors (TFTs). The in-plane hole mobility was measured in a bottom-gate, top-contact field effect transistor configuration. The hole mobility in these TFTs is higher parallel to the shearing direction than perpendicular.

Although the in-plane unit cell geometry becomes increasingly oblique as shearing speed increases (Fig. 4a and Supplementary Table 1), the average mobility along the shearing direction as a function of shearing speed increased to a maximum of $2.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for thin films prepared at a shearing speed of 2.6 mm s^{-1} (Fig. 4b and Supplementary Table 3). The average mobility decreased for faster shearing speeds, to a value of $0.47 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for samples prepared at a speed of 8 mm s^{-1} . We attribute the decrease in the average mobility observed at shearing speeds above 2.6 mm s^{-1} to the decrease in crystallite size and alignment (Fig. 1), causing an increase in the number of grain

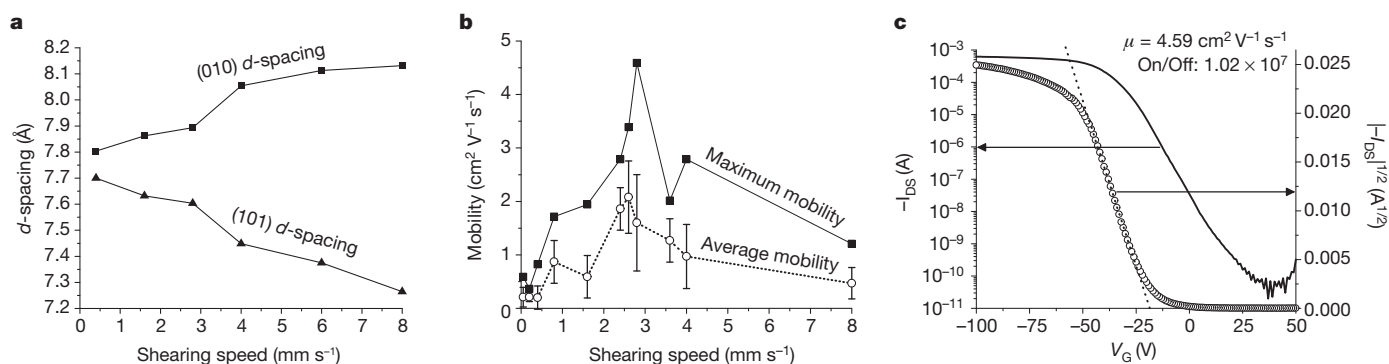


Figure 4 | Lattice strain and charge transport properties as a function of shearing speed. **a**, (101) and (010) d -spacing of TIPS-pentacene thin films sheared at different speeds. **b**, Average and maximum charge carrier mobilities of transistors fabricated from TIPS-pentacene thin films sheared at different speeds. The mobilities were measured along the direction of shearing. The error bars show the standard error of the mean, calculated by dividing the standard

deviation of the mobility by the square root of the total number of samples measured. The total number of samples tested is provided in Supplementary Table 3. **c**, Transfer curve of a device prepared at a shearing speed of 2.8 mm s^{-1} showing mobility of $4.6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. I_{DS} is the source-drain current. 'On/Off' represents the ratio between the maximum current in the device when it is active, and the minimum current when the device is turned off.

boundaries and contribution from the weaker electronic overlap of the T2 molecular pair respectively²⁶. The best-performing device showed a mobility of $4.6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, reached at a shearing speed of 2.8 mm s^{-1} (Fig. 4c). Mobility distributions for high-performance TFTs are shown in Supplementary Fig. 10. These TFTs were also stable for at least a year of storage in the dark and under vacuum, showing only a 10% decrease in mobility; additionally, no relaxation of lattice strain was observed for a film stored in air for more than three months (Supplementary Table 4). As a reference, the previously reported maximum mobility for a TIPS-pentacene TFT at room temperature was $1.8 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, and the observed threshold voltages here may be different from those reported previously owing to differences in device structure, material purity and interface properties^{21,25,27}. We attribute our higher mobility to the improved electronic coupling between molecules in the strained films. This conclusion is further supported by the observation that decreased mobilities were observed for partially strain-relieved films while maintaining the same film textures. The strained films were annealed with toluene vapour for 5 min, which resulted in a partial relief of the strain, as confirmed via GIXD (Supplementary Fig. 11). Longer exposure times formed cracks in the thin film. A decrease in TFT mobility after toluene vapour annealing was observed for all films prepared at speeds above 0.4 mm s^{-1} (Supplementary Fig. 12). There is no change in the film texture as confirmed by cross-polarized optical microscopy and atomic force microscopy, so we attribute the decreased mobility to relaxed molecular packing strain (Supplementary Fig. 11).

During solution shearing, the liquid film thickness decreases as a function of increasing shearing speed²⁴. A steeper temperature gradient forms in thinner liquid films, because the bottom of the film is in contact with a heating source, while the top surface is exposed to ambient temperature. The solvent evaporates faster in the thinner liquid film, which results in faster solution flow towards the growing crystal front (Supplementary Fig. 1). The faster solvent evaporation rate also gives the growing crystal front less time to reach an equilibrium state. We hypothesize that the fast crystallization and solvent evaporation kinetically traps metastable states. These states can then be relaxed to lower energy states using toluene vapour annealing. Faster rates of solvent evaporation coupled with velocity and concentration gradients are probably the key factors giving rise to strained crystals. We found that solution shearing can introduce lattice strain and charge transport enhancement in other organic semiconductors, demonstrating the utility of solution shearing as a general technique for tuning molecular packing (Supplementary Fig. 13).

In summary, we have demonstrated that solution shearing can be used to incrementally change the molecular packing of the TIPS-pentacene molecule. The π - π stacking distance of the T1 molecular pair decreased from 3.33 \AA to 3.08 \AA with increasing shearing speed, and density functional theory calculations show an increase in the charge transfer integral beyond that calculated for the evaporated thin film. The direction of elongated crystallite texture growth is along the b axis, which is also the direction for faster charge transport. We were able to achieve a maximum mobility as high as $4.6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, which is more than twice the highest maximum value ($1.8 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) previously reported²⁷. However, we did not observe a monotonic increase in mobility as a function of increasing shearing speed owing to the formation of smaller and less oriented crystallites at high shearing speeds, which introduce grain boundary regions and domains with weaker electronic overlap^{28,29}. Additionally, molecular displacements along the molecular long and short axes affect the charge transfer integral, so the overall charge transfer integral for all organic semiconductors may not necessarily increase with decreasing π - π stacking distance. The advantage of our solution-shearing approach is that it allows us to access many possible new packing structures that are not possible through other commonly used methods. Therefore, in a given molecular system it provides more opportunities to enhance the charge transport. Modifying the molecular packing by tuning the

extent of lattice strain instead of changing the chemical structure provides an alternative method of studying charge transport as a function of packing. We expect this work to provide a new way to improve the charge carrier mobility of organic semiconductors for practical applications.

METHODS SUMMARY

The TIPS-pentacene (received from 3M) was used without further purification. The bottom device substrate ($2 \times 2 \text{ cm}^2$) was treated with phenyltrichlorosilane (PTS) to ensure proper wetting for the TIPS-pentacene solution and to reduce surface charge traps. PTS treatment was accomplished by immersing the Piranha-cleaned (see the online-only Methods) wafer into a toluene solution of PTS (3 wt%) and heated at 90°C for 15 h. For the top shearing plate, a silicon wafer with 300 nm thermally grown oxide was treated with a monolayer of octadecyltrichlorosilane (OTS), according to a method we published previously for forming a highly hydrophobic crystalline OTS monolayer^{22,30}. The solution-sheared films were characterized using a cross-polarized optical microscope (Leica DM4000M). Thickness measurements were performed with a Dektak 150 profilometer, purchased from the Veeco Metrology Group. Tapping mode atomic force microscopy (AFM) was performed using a Multimode Nanoscope III (Digital Instruments/Veeco Metrology Group).

GIXD was performed at the Stanford Synchrotron Radiation Lightsource on beamlines 2-1, 7-2, and 11-3. Numerical integration of the diffraction peak areas was performed with the software WxDiff⁹. The unit cell geometries were determined by least-square-error optimization fitting of the peak positions. The molecular arrangements of TIPS-pentacene in the unit cells were obtained from a crystallographic refinement procedure. The electrical characterization was performed on bottom-gate, top-contact field effect transistors. All field effect transistors were fabricated and tested in ambient conditions, with exposure to light and air. The electrical characteristics of the TIPS-pentacene films were measured using a Keithley 4200-SCS semiconductor parameter analyser. The transfer integrals for the nearest-neighbour pairs of molecules (Supplementary Fig. 8) are calculated using density functional theory employing the B3LYP functional and the 6-31G(d,p) basis set.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 July; accepted 26 October 2011.

1. Dimitrakopoulos, C. D. & Malenfant, P. R. L. Organic thin film transistors for large area electronics. *Adv. Mater.* **14**, 99–117 (2002).
2. Jones, B. A. *et al.* High-mobility air-stable n-type semiconductors with processing versatility: dicyanoperylene-3,4,9,10-bis(dicarboximides). *Angew. Chem.* **116**, 6523–6526 (2004).
3. Rogers, J. A. *et al.* Paper-like electronic displays: large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. *Proc. Natl Acad. Sci. USA* **98**, 4835–4840 (2001).
4. Gao, P. *et al.* Dithieno[2,3-d;2',3'-d']benzo[1,2-b;4,5-b']dithiophene (DTBDT) as semiconductor for high-performance, solution-processed organic field-effect transistors. *Adv. Mater.* **21**, 213–216 (2009).
5. Zhang, M. *et al.* Field-effect transistors based on a benzothiadiazole–cyclopentadithiophene copolymer. *J. Am. Chem. Soc.* **129**, 3472–3473 (2007).
6. Lee, M. L., Fitzgerald, E. A., Bulsara, M. T., Currie, M. T. & Lochtefeld, A. Strained Si, SiGe, and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors. *J. Appl. Phys.* **97**, 011101 (2005).
7. Bredas, J. L., Calbert, J. P., da Silva Filho, D. A. & Cornil, J. Organic semiconductors: a theoretical characterization of the basic parameters governing charge transport. *Proc. Natl Acad. Sci. USA* **99**, 5804–5809 (2002).
8. Che, C.-M. *et al.* Single microcrystals of organoplatinum(II) complexes with high charge-carrier mobility. *Chem. Sci.* **2**, 216–220 (2011).
9. Chui, S. S. *et al.* Homoleptic platinum(II) and palladium(II) organothiolates and phenylselenolates: solvothermal synthesis, structural determination, optical properties, and single-source precursors for PdSe and PdS nanocrystals. *Chem. Asian J.* **5**, 2062–2074 (2010).
10. Sokolov, A. N. *Molecular Co-crystals: Semiconductors, Photoactive Solids, and Catalysts* PhD thesis, Univ. Iowa (2007).
11. Ni, Z. H. *et al.* Uniaxial strain on graphene: Raman spectroscopy study and band-gap opening. *ACS Nano* **2**, 2301–2305 (2008).
12. Anthony, J. E., Eaton, D. L. & Parkin, S. R. A road map to stable, soluble, easily crystallized pentacene derivatives. *Org. Lett.* **4**, 15–18 (2002).
13. Tang, M. L., Oh, J. H., Reichardt, A. D. & Bao, Z. Chlorination: a general route toward electron transport in organic semiconductors. *J. Am. Chem. Soc.* **131**, 3733–3740 (2009).
14. Okamoto, T. *et al.* Synthesis, characterization, and field-effect transistor performance of pentacene derivatives. *Adv. Mater.* **19**, 3381–3384 (2007).
15. Feng, X. *et al.* Towards high charge-carrier mobilities by rational design of the shape and periphery of discotics. *Nature Mater.* **8**, 421–426 (2009).

16. Gsänger, M. *et al.* A crystal-engineered hydrogen-bonded octachloroperylene diimide with a twisted core: an n-channel organic semiconductor. *Angew. Chem.* **122**, 752–755 (2010).
17. Chen, J., Anthony, J. & Martin, D. C. Thermally induced solid-state phase transition of bis(triisopropylsilyl)ethynyl pentacene crystals. *J. Phys. Chem. B* **110**, 16397–16403 (2006).
18. Yuan, Q. *et al.* Thin film structure of tetraceno[2,3-b]thiophene characterized by grazing incidence X-ray scattering and near-edge X-ray absorption fine structure analysis. *J. Am. Chem. Soc.* **130**, 3502–3508 (2008).
19. Mannsfeld, S. C. B., Tang, M. L. & Bao, Z. Thin film structure of triisopropylsilyl-ethynyl-functionalized pentacene and tetraceno[2,3-b]thiophene from grazing incidence X-ray diffraction. *Adv. Mater.* **23**, 127–131 (2011).
20. Sakanoue, T. & Sirringhaus, H. Band-like temperature dependence of mobility in a solution-processed organic semiconductor. *Nature Mater.* **9**, 736–740 (2010).
21. Lee, W. H. *et al.* Solution-processable pentacene microcrystal arrays for high performance organic field-effect transistors. *Appl. Phys. Lett.* **90**, 132106 (2007).
22. Becerril, H. A., Roberts, M. E., Liu, Z., Locklin, J. & Bao, Z. High-performance organic thin-film transistors through solution-sheared deposition of small-molecule organic semiconductors. *Adv. Mater.* **20**, 2588–2594 (2008).
23. Lovinger, A. J. & Wang, T. T. Investigation of the properties of directionally solidified poly(vinylidene fluoride). *Polymer* **20**, 725–732 (1979).
24. Rogowski, R. Z. & Darhuber, A. A. Crystal growth near moving contact lines on homogeneous and chemically patterned surfaces. *Langmuir* **26**, 11485–11493 (2010).
25. Sele, C. W. *et al.* Controlled deposition of highly ordered soluble acene thin films: effect of morphology and crystal orientation on transistor performance. *Adv. Mater.* **21**, 4926–4931 (2009).
26. Rivnay, J. *et al.* Large modulation of carrier transport by grain-boundary molecular packing and microstructure in organic thin films. *Nature Mater.* **8**, 952–958 (2009).
27. Park, S. K., Jackson, T. N., Anthony, J. E. & Mourey, D. A. High mobility solution processed 6,13-bis(triisopropyl-silyl)ethynyl pentacene organic thin film transistors. *Appl. Phys. Lett.* **91**, 063514 (2007).
28. Chen, J., Tee, C. K., Shtein, M., Anthony, J. & Martin, D. C. Grain-boundary-limited charge transport in solution-processed 6,13 bis(tri-isopropylsilyl)ethynyl pentacene thin film transistors. *J. Appl. Phys.* **103**, 114512–114513 (2008).
29. Lee, S. S. *et al.* Controlling nucleation and crystallization in solution-processed organic semiconductors for thin-film transistors. *Adv. Mater.* **21**, 3605–3609 (2009).
30. Ito, Y. *et al.* Crystalline ultrasoft self-assembled monolayers of alkylsilanes for organic field-effect transistors. *J. Am. Chem. Soc.* **131**, 9396–9404 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Salleo, J. E. Anthony, H. Li, A. Sokolov and J. Rivnay for discussions. We thank J. E. Anthony and M. M. Nelson of 3M Corp. for providing high-purity TIPS-pentacene. This publication was partially supported by the National Science Foundation DMR-Solid State Chemistry (DMR-0705687-002), the Samsung Advanced Institute of Technology, the Global Climate and Energy Project at Stanford University (SPO 25591130-45282-A) and the Air Force Office of Scientific Research (award number FA9550-09-1-0256). E.V. thanks the Eastman Kodak Corporation Kodak Fellows Program for support. Z.B. acknowledges support from the David Filo and Jerry Yang Faculty Fellowship from Stanford University. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of Stanford University, the Sponsors of the Global Climate and Energy Project, or others involved with the Global Climate and Energy Project.

Author Contributions G.G. and H.A.B. built the current version of the solution-shearing set-up. G.G. and E.V. performed X-ray and transistor measurements. S.C.B.M. performed unit cell and molecular packing calculations. G.G., E.V., S.C.B.M., D.H.K., M.F.T. and Z.B. analysed the X-ray data. S.A.-E. and A.A.-G. performed transfer integral calculations. G.G., E.V. and Z.B. wrote the manuscript, and all other authors had input. Z.B. and S.Y.L. directed the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Z.B. (zbao@stanford.edu).

METHODS

Materials. The TIPS-pentacene (received from 3M) was used without further purification. PTS and OTS were purchased from Sigma-Aldrich and used as received (storage under an argon atmosphere to prevent hydrolysis). Highly doped *n*-type silicon wafers (resistivity $<0.005 \Omega \text{ cm}$) with a 300-nm thermally grown silicon oxide gate dielectric film (capacitance of the gate dielectric per unit area, $C_{\text{ox}} = 10 \text{ nF cm}^{-2}$) were used as the substrates for TFT fabrication and GIXD.

Substrate preparation and characterization. The silicon wafers were cleaned in a Piranha solution (70/30 vol./vol. $\text{H}_2\text{SO}_4/\text{H}_2\text{O}_2$; a highly oxidative solution) for 25 min. The bottom device substrate ($2 \times 2 \text{ cm}^2$) was treated with PTS to ensure proper wetting for the TIPS-pentacene solution and to reduce surface charge traps. PTS treatment was accomplished by immersing the Piranha-cleaned wafer into a toluene solution of PTS (3 wt%) and heated at 90°C for 15 h. The substrate was subsequently removed from PTS solution and was sonicated for 2 min in toluene. It was then gently wiped with a sponge tip and rinsed sequentially with toluene, acetone and isopropanol. The water contact angle of the PTS treated surface ranged from 72° to 74° . The typical roughness (root mean square) of the surface was 0.3–0.5 nm.

For the top shearing plate, a silicon wafer with 300-nm thermally grown oxide was treated with a monolayer of OTS according to a method we published previously for forming a highly hydrophobic crystalline OTS monolayer^{22,30}. The OTS treatment is important to ensure deposition of TIPS-pentacene only on the bottom PTS-treated substrate. The water contact angle of the OTS modified surface ranged between 102° and 104° , and the roughness of the surface ranged between the root-mean-square value of 0.2 nm and 0.5 nm.

Solution shearing of TIPS-pentacene films. The TIPS-pentacene solution was prepared at a concentration of 8 mg ml^{-1} in toluene. Both the device substrate and the shearing plate were held in place by vacuum while the device substrate was mounted on a resistively heated stage (a thermoelectric module from Custom Thermoelectric) held at 90°C . After placing about $40 \mu\text{l cm}^{-2}$ of TIPS-pentacene solution on the device substrate, the shearing plate was lowered with a micromanipulator to make contact with the solution. The device substrate was horizontal while the shearing plate was placed at a tilt angle of 8° from the horizontal. The gap distance (d , shown in Supplementary Fig. 1) between the device substrate and the shearing plate was fixed at $100 \mu\text{m}$. The shearing plate was translated at different velocities by a stepper motor. The resulting sheared film was left on the heating stage for 2–3 min at 90°C to remove residual solvent.

Solution-sheared film characterizations. The solution-sheared films were characterized using a cross-polarized optical microscope (Leica DM4000M).

Thickness measurements were performed with a Dektak 150 profilometer, purchased from the Veeco Metrology Group. Tapping-mode atomic force microscopy was performed using a Multimode Nanoscope III (Digital Instruments/Veeco Metrology Group).

GIXD was performed at the Stanford Synchrotron Radiation Lightsource on beamlines 2-1, 7-2 and 11-3. On the 11-3 beamline, a photon energy of 12.73 keV was used, and the scattering intensity was recorded on a two-dimensional image plate (MAR-345) with a pixel size of $150 \mu\text{m}$ ($2,300 \times 2,300$ pixels), located at a distance of 400 mm from the sample centre. The distance between the sample and the detector was calibrated using a LaB_6 polycrystalline standard. The incidence angle was chosen in the range 0.10° – 0.12° . The beam size was $50 \mu\text{m} \times 150 \mu\text{m}$, which resulted in a beam exposure on the sample $150 \mu\text{m}$ wide over the entire length of the 10-mm sample. The data was distortion-corrected (theta-dependent image distortion introduced by planar detector surface) before performing quantitative analysis on the images. Numerical integration of the diffraction peak areas was performed with the software WxDiff⁹. The overall resolution in the GIXD experiments, dominated by the sample size, was about 0.01 \AA^{-1} . The photon energy on beamlines 2-1 and 7-2 was 8 keV. Beamline 2-1 was used for high-resolution specular scanning with a point detector. Analysis software included Origin (from Microcal) and GSAS (Rietveld analysis). Beamline 7-2 was used for high-resolution grazing incidence scans, and uses a Huber high-resolution diffractometer.

The electrical characterization was performed on bottom-gate, top-contact field effect transistors created by thermally evaporating 40-nm gold source and drain electrodes on the solution-sheared samples through a shadow mask. All field effect transistors were fabricated and tested in ambient conditions, with exposure to light and air. The devices were stored in the dark under vacuum. The electrical characteristics of the TIPS-pentacene films were measured using a Keithley 4200-SCS semiconductor parameter analyser. Transfer characteristics (I_{DS} as a function of V_{G} , where I_{DS} is the source–drain current and V_{G} is the gate voltage) were tested for each transistor, and output characteristics (I_{DS} as a function of V_{D} at various gate voltages, where V_{D} is the drain voltage) were measured for representative transistors. The $I_{\text{DS}}-V_{\text{D}}$ curves were collected with V_{G} increasing from -100 V to 0 V and with V_{D} sweeping from 0 V to -100 V . The $I_{\text{DS}}-V_{\text{G}}$ curves were collected with V_{G} decreasing from 50 V to -100 V at a constant V_{D} of -100 V . The saturation mobility μ was extracted from the slope of the transfer curve $V_{\text{G}}-\sqrt{I_{\text{DS}}}$, where $I_{\text{DS}} = \mu WC_{\text{ox}}(V_{\text{G}} - V_{\text{T}})^2/2L$, where W ($=1,000 \mu\text{m}$) and L ($=50 \mu\text{m}$) are the channel width and length, respectively, C_{ox} is the capacitance of the gate dielectric per unit area, and V_{T} is the threshold voltage.

Forcing of wet phases in southeast Africa over the past 17,000 years

Enno Schefuß¹, Holger Kuhlmann¹, Gesine Mollenhauer^{1,2}, Matthias Prange¹ & Jürgen Pätzold¹

Intense debate persists about the climatic mechanisms governing hydrologic changes in tropical and subtropical southeast Africa since the Last Glacial Maximum, about 20,000 years ago. In particular, the relative importance of atmospheric and oceanic processes is not firmly established^{1–5}. Southward shifts of the intertropical convergence zone (ITCZ) driven by high-latitude climate changes have been suggested as a primary forcing^{2,3}, whereas other studies infer a predominant influence of Indian Ocean sea surface temperatures on regional rainfall changes^{4,5}. To address this question, a continuous record representing an integrated signal of regional climate variability is required, but has until now been missing. Here we show that remote atmospheric forcing by cold events in the northern high latitudes appears to have been the main driver of hydro-climatology in southeast Africa during rapid climate changes over the past 17,000 years. Our results are based on a reconstruction of precipitation and river discharge changes, as recorded in a marine sediment core off the mouth of the Zambezi River, near the southern boundary of the modern seasonal ITCZ migration. Indian Ocean sea surface temperatures did not exert a primary control over southeast African hydrologic variability. Instead, phases of high precipitation and terrestrial discharge occurred when the ITCZ was forced southwards during Northern Hemisphere cold events, such as Heinrich stadial 1 (around 16,000 years ago) and the Younger Dryas (around 12,000 years ago), or when local summer insolation was high in the late Holocene, that is, during the past 4,000 years.

Climate changes reconstructed from sediments in Lake Malawi, the southernmost of the East African great lakes and located within the Zambezi catchment, point to arid conditions and strong northerly wind anomalies during Northern Hemisphere cold events, such as the Younger Dryas^{2,3}. It has been inferred that these periods represent southward shifts of the ITCZ^{2,3,6}. Dry conditions during the Younger Dryas and Heinrich stadial 1 (HS1) in Lake Tanganyika, located closer to the Equator in East Africa, were interpreted to have been caused by lowered Indian Ocean sea surface temperature⁴ (SST). Recently, widespread drought conditions during HS1 in East Africa have similarly been suggested to be driven by cold conditions in the Indian Ocean⁵. Such interpretations are based on meteorological observations that modern rainfall in eastern and southern Africa is strongly related to high SST in the western and southwestern Indian Ocean, respectively⁷. In contrast, Lake Chilwa—located southeast of Lake Malawi—recorded high-stands, which appear to be solely associated with Northern Hemisphere cold events⁸. Further westward in the interior of subtropical southern Africa, palaeo-environmental information is sparse. Age dating of dunes in western Zambia and western Zimbabwe points to dry conditions at 18,000–17,000, 15,000–14,000, 11,000–8,000 and 6,000–4,000 years before present (yr BP; ref. 9), implying that non-dune-building periods around 16,000 yr BP and from 13,000 to 12,000 yr BP roughly correspond to the wet periods found at Lake Chilwa⁸. Palaeo-shorelines and other geomorphologic evidence from the Central Kalahari point to the existence of an extended lake system

around the time of HS1 and around 8,500 yr BP, with evidence for the existence of smaller lakes around the time of the Younger Dryas¹⁰. It thus appears that the climatic history of subtropical southern Africa is complex, and the importance of the various potential climatic forcing mechanisms remains unresolved.

To provide more insights into the climatic forcing of hydrology and land–ocean linkages in southeast Africa, we present a reconstruction of precipitation changes in the Zambezi catchment and its sedimentary discharge fluctuations in conjunction with SST estimates of the southwest Indian Ocean. The Zambezi catchment is located at the southern boundary of the present-day seasonal ITCZ migration, and is thus ideally suited to have recorded past changes in its southward extension. Under modern conditions, austral summer rainfall associated with the ITCZ extends to 15–20° S (ref. 6; Fig. 1a, b). A low-pressure cell forms over southern Africa during summer; it attracts

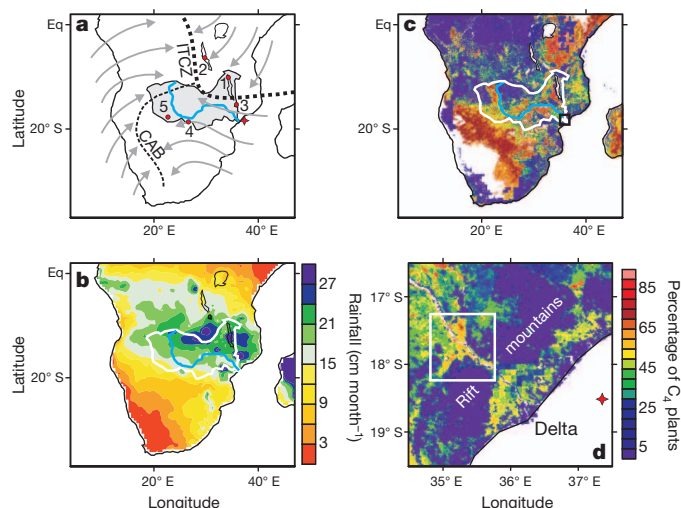


Figure 1 | Modern atmospheric circulation over southern Africa during Southern Hemisphere summer, summer rainfall and vegetation types. **a**, Schematic representation of the atmospheric circulation over southern Africa during austral summer (December, January, February; DJF) with approximate position of the ITCZ and the Congo Air Boundary (CAB). Indicated are: the main course of the Zambezi River (blue), the Zambezi catchment (grey) and the location of Geob9307-3 (red diamond). Filled red circles indicate locations of other records discussed in the text: (1) Lake Malawi², (2) Lake Tanganyika⁴, (3) Lake Chilwa⁸, (4) dunes in Zambia and Zimbabwe⁹, and (5) palaeo-shorelines in the Kalahari¹⁰. **b**, Austral summer (DJF) rainfall for the period 1950–99 (University of Delaware data set; http://www.esrl.noaa.gov/psd/data/gridded/data.UDEL_AirT_Precip.html). Outline of the Zambezi catchment in white and main course of the Zambezi River in blue. **c**, Modern-day vegetation type distribution¹⁵, shown as percentage of C₄ plants. White areas are either >95% C₄ or not vegetated. Zambezi catchment and Zambezi River as in **b**. Colour scale as in **d**. Black box indicates the area shown in **d**. **d**, Map of the relative C₄ plant abundance¹⁵ in the lower Zambezi catchment. White areas are not vegetated. White box indicates the location of the lower Zambezi floodplain.

¹MARUM – Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, D-28359 Bremen, Germany. ²Alfred Wegener Institute for Polar and Marine Research, D-27568 Bremerhaven, Germany.

moisture from the tropical Atlantic by tropical westerly air flow, and from the Indian Ocean by easterly air flow⁶. Atlantic moisture influences are restricted to the highlands south of the Congo basin, separated by the Congo Air Boundary from the majority of subtropical southern Africa that receives moisture from the Indian Ocean¹¹. During austral winter, a high-pressure system over southern Africa leads to dry conditions, except in a small winter rainfall zone at its southwestern tip⁶. The Zambezi River is the fourth-largest river in Africa, originating in Zambia and flowing southeastward to the Indian Ocean. The upper Zambezi is separated from its lower part by the Victoria Falls. After flowing through a series of gorges, the river enters a broad valley and spreads out over a large floodplain in its lower part. About 150 km from the coast, the catchment narrows where the river flows through the eastern Rift mountains. Maximum rainfall in the catchment during the peak of the southward ITCZ migration occurs in this area¹² (Fig. 1b), periodically causing extensive flooding¹³. The floodplain is dominated by papyrus swamps (dominated by *Cyperus papyrus*, a C₄ plant) with a lower importance of reed swamps (*Phragmites* and *Typha*, C₃ plants)¹⁴. The C₄ plant dominance in the lower Zambezi floodplain is seen in the relative abundance of C₄ plants¹⁵ (Fig. 1c, d). Downstream from the Rift mountains, the Zambezi splits up into a flat and wide delta¹³. Woodlands, savannah and extensive mangrove forests (C₃ plants) vegetate the delta and coastal areas¹⁴.

The 6.51-m-long marine sediment core GeoB9307-3 (18° 33.9' S, 37° 22.8' E, 542 m water depth) was retrieved about 100 km off the Zambezi delta. The core location is in the zone of high deglacial-Holocene sedimentation of material discharged by the Zambezi River¹⁶. The chronology of the core is established by 19 ¹⁴C AMS (accelerator mass spectrometry) dates on mixed planktonic foraminifera (Supplementary Table 1). On the basis of this chronology, the core covers the past 17,000 years. The average time resolution between the 125 samples is 130 years. For the time periods of HS1 and the Younger Dryas, we detect large increases in sedimentary terrestrial versus marine elemental ratios (Fig. 2b), in relative soil organic matter contributions¹⁷ (branched and isoprenoid tetraether (BIT) index, Fig. 2c), and in sedimentary concentrations of terrestrial-plant-derived long-chain *n*-alkanes (Fig. 2d), indicating enhanced terrestrial discharge by the Zambezi River. The elevated concentrations of *n*-alkanes indicate sustained vegetation cover during HS1 and the Younger Dryas, and decreases of soil pH estimates during these intervals (Supplementary Fig. 4) suggest higher rainfall as primary cause. For all these parameters, a slight increase is also detected for the late Holocene compared to the early Holocene. Comparison to records from Lake Malawi^{2,3} suggests that the rainfall maximum associated with the ITCZ shifted southward of the Malawi basin during HS1 and the Younger Dryas, causing arid conditions at Lake Malawi while increasing rainfall in the lower Zambezi catchment. The stronger increases in these discharge-related parameters during the Younger Dryas and HS1 compared to the late Holocene points to the influence of sea-level changes. Sea level was reduced during the last deglaciation¹⁸, resulting in a more proximal location of the Zambezi outflow to the core site.

In order to investigate the continental hydrologic changes without the influence of sea-level changes, we analysed the hydrogen isotope composition of lipid biomarkers derived from higher plants. Hydrogen isotope compositions of the *n*-C₃₁ alkane range from −136‰ to −164‰ (relative to the VSMOW standard), showing depleted values from 16,950 to 14,700 yr BP, corresponding to the period of HS1, and from 12,800 to 11,350 yr BP, representing the Younger Dryas period (Fig. 2e). The most enriched values are found in the early Holocene, followed by a gradual decline towards the late Holocene that is interrupted by several short-term excursions to more enriched values. The *n*-C₃₁ alkane derives from the protective wax coating of leaves of terrestrial higher plants¹⁹. Its hydrogen isotope composition reflects changes in the isotopic composition of precipitation, with potential enrichment due to evaporation from soils and evapo-transpiration from leaves²⁰, which amplify the recorded signal

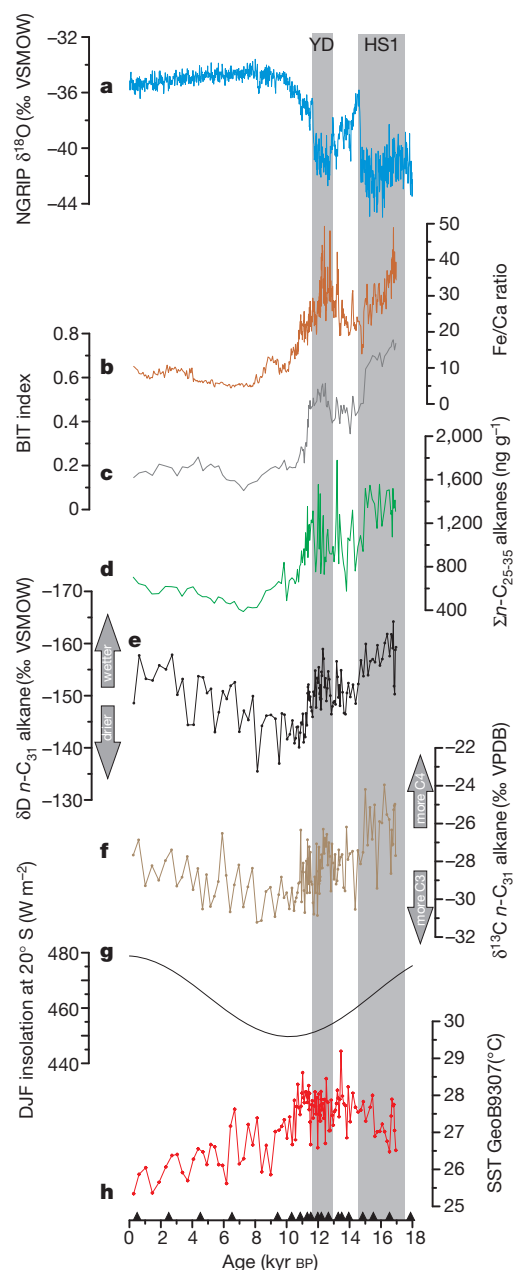


Figure 2 | Records of environmental variability in southeastern Africa over the past 17,000 years. **a**, Oxygen isotope changes in NGRIP ice core²⁸, mainly reflecting northern high-latitude temperature changes. **b**, Fe/Ca ratio, indicating terrestrial versus marine elemental contributions to sediments of GeoB9307-3. **c**, BIT index of GeoB9307-3, reflecting changes in relative soil organic matter contribution. **d**, Concentrations of long-chain, odd-numbered *n*-alkanes (*n*-C₂₅ to *n*-C₃₅) in the sediments of GeoB9307-3. **e**, Hydrogen isotope compositions of the *n*-C₃₁ alkane in GeoB9307-3, reflecting rainfall changes in the Zambezi catchment. Error bars based on replicate analyses are shown in Supplementary Fig. 1. **f**, Stable carbon isotope compositions of *n*-C₃₁ alkanes in GeoB9307-3, indicating relative contribution changes from the lower Zambezi floodplain. Error bars based on replicate analyses are shown in Supplementary Fig. 2. **g**, Long-term insolation changes over southern Africa (20° S) during Southern Hemisphere summer (DJF)²⁵. **h**, TEX₈₆-derived SST record from GeoB9307-3, reflecting ocean temperature changes in the southwest Indian Ocean. Vertical grey bars indicate intervals of Heinrich Stadial 1 (HS1) and the Younger Dryas (YD). Black triangles indicate ¹⁴C AMS ages.

under arid conditions²¹. In monsoonal regions, depleted isotopic compositions of precipitation are indicative of changes in the amount of rainfall²². These data suggest that changes in rainfall amount caused the discharge pulses detected for the Younger Dryas and HS1.

Comparison of the δD values of the long-chain plant-waxes with their stable carbon isotope ($\delta^{13}C$) compositions reveals that sedimentary n -alkanes with depleted δD values show a stronger C_4 plant signal and vice versa (Fig. 2f). As C_4 plants are usually more drought-tolerant than C_3 plants²³, this correlation is unexpected. Although long-term vegetation changes are not ruled out (see Supplementary Information), the strong coherence of both isotopic signals rather points to a shift in source area associated with the hydrologic changes as the primary cause for the observed relation. We infer that when maximum rainfall and flooding occurs in the floodplain, as under modern-day conditions, more C_4 -plant-derived waxes are exported to the ocean, as reflected in the core-top isotopic signature (Figs 1d, 2e). Under more arid conditions, such as during the early Holocene, the export of C_4 -plant-derived lipids from the floodplain was diminished, and only low amounts of predominantly C_3 -plant-derived waxes from near-coastal vegetation were exported. We thus deduce that the terrestrial signal in core GeoB9307-3 is predominantly derived from the lower Zambezi catchment, dominated by changes in relative contributions from the floodplain. Aeolian transport of terrigenous material from subtropical southern Africa towards the southwest Indian Ocean is negligible, owing to the prevailing easterly winds¹².

Part of the signal seen in the δD record may thus derive from the changes in relative contributions from different vegetation types, seen in the C_3/C_4 changes (see, for example, ref. 24). Although we cannot completely rule out this possible effect, the evidence provided by the other parameters (Fe/Ca ratio, BIT index, n -alkane concentrations, soil pH estimates) indicates that the depleted δD values during HS1, the Younger Dryas and the late Holocene indeed reflect increased precipitation in the lower Zambezi catchment.

The long-term trends in the δD and $\delta^{13}C$ records correlate with mean summer insolation (Fig. 2g) over southern Africa²⁵: higher insolation intensifies atmospheric convection, leading to higher rainfall over that region²⁶. For Heinrich events, modelling results indicate an increase of precipitation associated with isotopically depleted rainfall over southern Africa²⁷, consistent in magnitude with the isotopic changes detected for HS1. A detailed comparison of the plant-wax isotopic changes with the oxygen isotope changes in the NGRIP ice core²⁸ (Fig. 3), which reflect northern high-latitude temperature changes, reveals a synchronous onset and end of the Younger Dryas

and a synchronous end of HS1. This synchronicity supports our conclusion that the plant waxes are not transported over long distances. As suggested by other studies^{8–10}, rainfall during HS1 and the Younger Dryas probably also increased in the western parts of the Zambezi catchment. However, plant waxes derived from these more inland parts would carry a much more depleted δD signal due to increased moisture rainout during long-distance transport²⁹. We thus infer that long-distance transport of plant waxes by the Zambezi River is of minor importance. Furthermore, the detected synchronicity of isotopic changes suggests a direct forcing of southeast African wet phases by Northern Hemisphere cold events, and points to a large-scale atmospheric teleconnection (see, for example, ref. 30). It is likely that persistent cold conditions in the Northern Hemisphere forced the ITCZ to a more southerly position during these events. The deviations in the amplitudes of isotopic changes of the ice core and the plant waxes during HS1 (Fig. 3) are probably explained by the modulation of rainfall intensity by local insolation.

In order to evaluate the effect of southwest Indian Ocean SST changes on continental hydrology, we compare the terrestrial records to a TEX₈₆-based SST record from GeoB9307-3 (Fig. 2h). The TEX₈₆ index reflects distributional variations in membrane lipids from marine Thaumarchaeota, related to ocean temperature (see Supplementary Information). The high BIT values in part of the record did not lead to a substantial bias in SST estimates (see Supplementary Information). The SST record shows warm conditions during the Younger Dryas, but does not indicate similarly warm conditions during HS1. Moreover, the timing of SST changes does not correspond to the observed hydrologic changes. Therefore, we suggest that sufficiently warm conditions in the southwest Indian Ocean may have been an important pre-condition for southward movements of the ITCZ and the supply of moisture, but that rainfall changes in the Zambezi catchment over these timescales were not primarily driven by these SSTs. We find no evidence for drought conditions in southeast Africa during HS1 driven by cold conditions in the Indian Ocean⁵.

In summary, we conclude that rainfall and discharge changes in the Zambezi catchment result from a combination of local insolation and latitudinal ITCZ shifts, with the latter being directly forced by high-latitude climate changes in the Northern Hemisphere. SST changes in the southwest Indian Ocean appear not to be the primary forcing of past hydrologic changes, but may be an important prerequisite for such changes.

METHODS SUMMARY

Element intensities were measured at 1 cm resolution using a XRF-I core scanner at MARUM, University of Bremen. The central sensor unit is equipped with a molybdenum X-ray source (3–50 keV), a Peltier-cooled PSI detector with a 125 μm beryllium window, and a multichannel analyser with 20 eV spectral resolution. XRF (X-ray fluorescence) data for this study were collected over a 1 cm² area using 30 s count time, and conditions of 20 kV and 0.087 mA for the X-ray source.

Sediments were extracted with a Dionex accelerated solvent extractor using dichloromethane:methanol (9:1; v/v). Saturated hydrocarbon fractions were obtained using silica column chromatography by elution with hexane and subsequent elution over AgNO₃-coated silica. Squalane was added before extraction as an internal standard. Hydrogen isotope compositions of lipids were analysed on a Trace gas chromatograph coupled via a pyrolysis reactor to a MAT 253 mass spectrometer. Isotope values were measured against calibrated H₂ reference gas. δD values are reported in ‰ versus VSMOW. Further technical information is provided in Supplementary Information.

Compound-specific stable carbon isotope values were analysed on a Trace gas chromatograph coupled via a combustion reactor to a MAT 252 mass spectrometer. Isotope values were measured against calibrated CO₂ reference gas. $\delta^{13}C$ values are reported in ‰ versus VPDB. Further technical information is provided in Supplementary Information.

Polar fractions of the total lipid extracts eluted from silica columns with dichloromethane:methanol (1:1; v/v) were filtered through a 4- μm pore size PTFE filter and dissolved in hexane/isopropanol (99:1; v/v) before analysis by high performance liquid chromatography/atmospheric pressure chemical ionization–mass

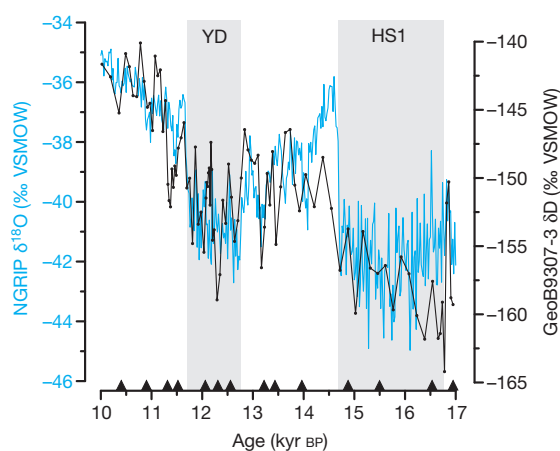


Figure 3 | Detailed comparison of Greenland climate changes with hydrologic variations in the Zambezi catchment during the last deglaciation. Greenland temperature changes reflected in oxygen isotope compositions (blue) of the NGRIP ice core²⁸ compared to hydrogen isotope changes of the n -C₃₁ alkane in sediments of GeoB9307-3 reflecting rainfall changes in the Zambezi catchment (black, note inverted scale). Deviations in amplitudes are partly due to modulation of rainfall intensity by local insolation (Fig. 2g). Grey bars and black triangles as in Fig. 2. Error bars based on replicate analyses are shown in Supplementary Fig. 1.

spectrometry (HPLC/APCI-MS). An Agilent 1200 series HPLC/APCI-MS system equipped with a Grace Prevail Cyano column (150 mm × 2.1 mm; 3 µm) was used, and separation was achieved in normal phase.

Received 11 May; accepted 27 October 2011.

- Clement, A. C. & Peterson, L. C. Mechanisms of abrupt climate change of the last glacial period. *Rev. Geophys.* **46**, RG4002 (2008).
- Johnson, T. C. *et al.* A high-resolution paleoclimate record spanning the past 25,000 years in southern east Africa. *Science* **296**, 113–132 (2002).
- Castañeda, I. S., Werne, J. P. & Johnson, T. C. Wet and arid phases in the southeast African tropics since the Last Glacial Maximum. *Geology* **35**, 823–826 (2007).
- Tierney, J. E. *et al.* Northern Hemisphere controls on tropical southeast African climate during the past 60,000 years. *Science* **322**, 252–255 (2008).
- Stager, C. J., Ryves, D. B., Chase, B. M. & Pausata, F. S. R. Catastrophic drought in the Afro-Asian monsoon region during Heinrich Event 1. *Science* **331**, 1299–1302 (2011).
- Gasse, F. *et al.* Climatic patterns in equatorial and southern Africa from 30,000 to 10,000 years ago reconstructed from terrestrial and near-shore proxy data. *Quat. Sci. Rev.* **27**, 2316–2340 (2008).
- Jury, M. R., Enfield, D. B. & Melice, J. L. Tropical monsoons around Africa: stability of El Niño–Southern Oscillation associations and links with continental climate. *J. Geophys. Res.* **107**(C10), 3151, <http://dx.doi.org/10.1029/2000JC000507> (2002).
- Thomas, D. S. G. *et al.* Late Quaternary highstands at Lake Chilwa, Malawi: frequency, timing and possible forcing mechanisms in the last 44 ka. *Quat. Sci. Rev.* **28**, 526–539 (2009).
- Thomas, D. S. G. & Shaw, P. A. Late Quaternary environmental change in central southern Africa: new data, synthesis, issues and prospects. *Quat. Sci. Rev.* **21**, 783–797 (2002).
- Burrough, S. L., Thomas, D. S. G. & Singarayer, J. S. Late Quaternary hydrological dynamics in the Middle Kalahari: forcing and feedbacks. *Earth Sci. Rev.* **96**, 313–326 (2009).
- Gimeno, L. *et al.* On the origin of continental precipitation. *Geophys. Res. Lett.* **37**, L13804, <http://dx.doi.org/10.1029/2010GL043712> (2010).
- Nicholson, S. E. The nature of rainfall variability over Africa on time scales of decades to millennia. *Glob. Planet. Change* **26**, 137–158 (2000).
- Beilfuss, R. & dos Santos, D. *Patterns of Hydrological Change in the Zambezi Delta, Mozambique* (Working Paper No. 2, International Crane Foundation, Sofala, Mozambique, 2001); available at (http://www.savingcranes.org/images/stories/pdf/conservation/Zambezi_hydrology_Working_Paper2.pdf).
- Beilfuss, R., Moore, D., Bento, C. & Dutton, P. *Patterns of Vegetation Change in the Zambezi Delta, Mozambique* (Working Paper No. 3, International Crane Foundation, Sofala, Mozambique, 2001); available at (http://www.savingcranes.org/images/stories/pdf/conservation/Zambezi_Vegetation_Working_Paper3.pdf).
- Still, C. J. & Powell, R. L. in *Understanding Movement, Pattern, and Process on Earth through Isotope Mapping* (eds West, J. B., Bowen, G. J., Dawson, T. E. & Tu, K. P.) 179–194 (Springer, 2010).
- Schulz, H., Lückge, A., Emeis, K.-C. & Mackensen, A. Variability of Holocene to Late Pleistocene Zambezi riverine sedimentation at the upper continental slope off Mozambique, 15°–21°S. *Mar. Geol.* **286**, 21–34 (2011).
- Hopmans, E. C., Schefuß, E., Sinninghe Damsté, J. S. & Schouten, S. A novel proxy for terrestrial carbon input based on branched and isoprenoid tetraethers. *Earth Planet. Sci. Lett.* **224**, 107–116 (2004).
- Fairbanks, R. G. Sea level rise during the last deglaciation as recorded in Barbados corals. *Nature* **342**, 637–642 (1989).
- Eglinton, G. & Hamilton, R. J. Leaf epicuticular waxes. *Science* **156**, 1322–1335 (1967).
- Sachse, D., Radke, J. & Gleixner, G. δD values of individual *n*-alkanes from terrestrial plants along a climatic gradient — implications for the sedimentary biomarker record. *Org. Geochem.* **37**, 469–483 (2006).
- Schefuß, E., Schouten, S. & Schneider, R. R. Climatic controls of Central African hydrology during the last 20,000 years. *Nature* **437**, 1003–1006 (2005).
- Risi, C., Bony, S. & Vimeux, F. Influence of convective processes on the isotopic composition (δ¹⁸O and δD) of precipitation and water vapor in the tropics: 2. Physical interpretation of the amount effect. *J. Geophys. Res.* **113**, D19306, <http://dx.doi.org/10.1029/2008JD009943> (2008).
- Ehleringer, J. R. & Cerling, T. E. in *Encyclopedia of Global Environmental Change* Vol. 2 (ed. Munn, T.) 186–190 (Wiley & Sons, 2002). pp. 186.
- Hou, J. Z., D'Andrea, W. J., MacDonald, D. & Huang, Y. S. Hydrogen isotopic variability in leaf waxes among terrestrial and aquatic plants around Blood Pond, Massachusetts (USA). *Org. Geochem.* **38**, 977–984 (2007).
- Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
- Partridge, T. C. *et al.* Orbital forcing of climate over South Africa: a 200,000 year rainfall record from the Pretoria Saltpan. *Quat. Sci. Rev.* **16**, 1125–1133 (1997).
- Lewis, S. C., LeGrande, A. N., Kelley, M. & Schmidt, G. A. Water vapour source impacts on oxygen isotope variability in tropical precipitation during Heinrich events. *Clim. Past* **6**, 325–343 (2010).
- Andersen, K. K. *et al.* High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**, 147–151 (2004).
- Frankenberg, C. *et al.* Dynamic processes governing lower-tropospheric HDO/H₂O ratios as observed from space and ground. *Science* **325**, 1374–1377 (2009).
- Broccoli, A. J., Dahl, K. A. & Stouffer, R. J. Response of the ITCZ to Northern Hemisphere cooling. *Geophys. Res. Lett.* **33**, L01702, <http://dx.doi.org/10.1029/2005GL024546> (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the DFG Research Center/Cluster of Excellence ‘The Ocean in the Earth System’ and DFG grant Sche903/8. Laboratory assistance was provided by K. Siedenberg, M. Segl, W. Bevern, B. Meyer-Schack and R. Kreutz. We thank R. Schneider for logistical support and S. Weldeab for sample preparation for radiocarbon analyses. Comments by Y. Huang and J. Russell improved the manuscript.

Author Contributions Experimental work was carried out by E.S., H.K., G.M. and J.P.; data analysis and interpretation were carried out by E.S., H.K., G.M., M.P. and J.P.

Author Information Data reported here are stored in the Pangaea database (<http://www.pangaea.de/10.1594/PANGAEA.771396>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.S. (schefuss@uni-bremen.de).

Lowland–upland migration of sauropod dinosaurs during the Late Jurassic epoch

Henry C. Fricke¹, Justin Henceroth¹ & Marie E. Hoerner^{1†}

Sauropod dinosaurs were the largest vertebrates ever to walk the Earth, and as mega-herbivores they were important parts of terrestrial ecosystems. In the Late Jurassic-aged Morrison depositional basin of western North America, these animals occupied lowland river-floodplain settings characterized by a seasonally dry climate^{1,2}. Massive herbivores with high nutritional and water needs could periodically experience nutritional and water stress under these conditions, and thus the common occurrence of sauropods in this basin has remained a paradox. Energetic arguments and mammalian analogues have been used to suggest that migration allowed sauropods access to food and water resources over a wide region or during times of drought or both^{3,4}, but there has been no direct support for these hypotheses. Here we compare oxygen isotope ratios ($\delta^{18}\text{O}$) of tooth-enamel carbonate from the sauropod *Camarasaurus* with those of ancient soil, lake and wetland (that is, ‘authigenic’) carbonates that formed in lowland settings. We demonstrate that certain populations of these animals did in fact undertake seasonal migrations of several hundred kilometres from lowland to upland environments. This ability to describe patterns of sauropod movement will help to elucidate the role that migration played in the ecology and evolution of gigantism of these and associated dinosaurs.

Inferring the behaviour of ancient organisms is difficult, but geochemical information preserved in their fossil remains can provide such an opportunity. This study of sauropod dinosaur behaviour relies on the fact that $\delta^{18}\text{O}$ values of surface waters ($\delta^{18}\text{O}_{\text{sf}}$ for example streams, lakes) vary significantly over any given landscape in response to differences in aridity and elevation among other environmental factors^{5,6}. Authigenic carbonates (CaCO_3) form in basin soils, lakes and wetlands, and record the oxygen isotopic characteristics of these host isotopic domains when they precipitate. Similarly vertebrate tooth enamel (bioapatite $\text{Ca}_5(\text{PO}_4)_3(\text{OH}, \text{CO}_3)$) records the oxygen isotope characteristics of the surface water reservoirs that serve as their drinking water^{7,8}. If $\delta^{18}\text{O}_{\text{sf}}$ inferred from ‘non-migratory’ authigenic carbonates and from dinosaur tooth enamel differ, then it can be concluded that dinosaurs were drinking water that fell outside the basin and thus they travelled outside it.

To use this approach we analysed enamel carbonate from teeth ($n = 32$) of *Camarasaurus* sp. and *Camarasaurus lentus* collected at Thermopolis, Wyoming, and Dinosaur National Monument, Utah (DNM), respectively (Fig. 1a). Palaeosol and lacustrine carbonates were also analysed from DNM ($n = 38$; see Supplementary Information for details on methods and statistics). In addition, we used published $\delta^{18}\text{O}$ data obtained from a variety of authigenic carbonates found over the entire Morrison basin including the Thermopolis area^{9–12}. Comparisons of isotopic data from co-occurring authigenic carbonates and tooth enamel, from tooth-enamel carbonate and tooth-enamel phosphate, and from single teeth indicate that primary palaeobiological information is preserved in tooth enamel (see Supplementary Information for more details about diagenesis).

To estimate $\delta^{18}\text{O}_{\text{sf}}$ using dinosaur tooth enamel, it is assumed that they fractionated oxygen isotopes in a manner similar to all water-dependent

vertebrates studied so far, including birds, mammals and reptiles^{7,8}. To estimate $\delta^{18}\text{O}_{\text{sf}}$ using authigenic carbonate, it is assumed that oxygen isotope fractionation occurred at 24 °C, a temperature consistent with modelled mean annual temperature for the region¹³ (see

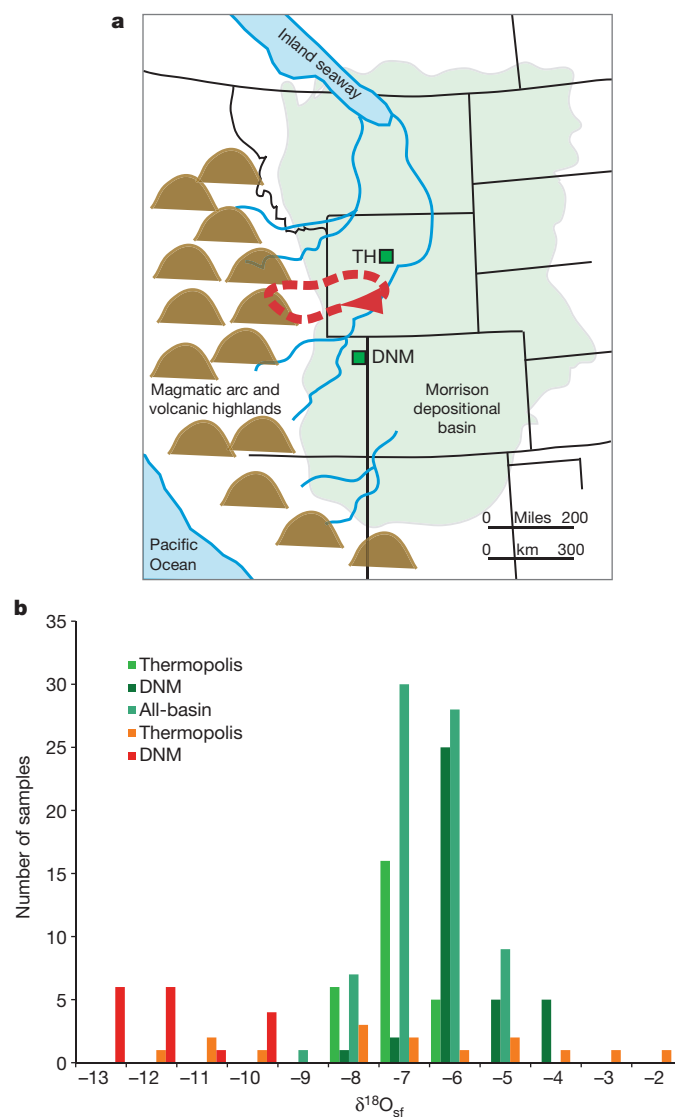


Figure 1 | Fossil localities, inferred oxygen isotope ratios of surface water and possible *Camarasaurus* migration routes. **a**, Palaeogeography of western North America during late Jurassic/Morrison time (after refs 2, 9), including fossil localities and one hypothetical migration route. **b**, $\delta^{18}\text{O}_{\text{sf}}$ estimated using tooth enamel (reds) and authigenic carbonates (greens; Thermopolis data from ref. 11; all-basin data from refs 9, 10, 12). See text and Supplementary Information for details.

¹Department of Geology, Colorado College, Colorado Springs, Colorado 80903, USA. [†]Present address: Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA.

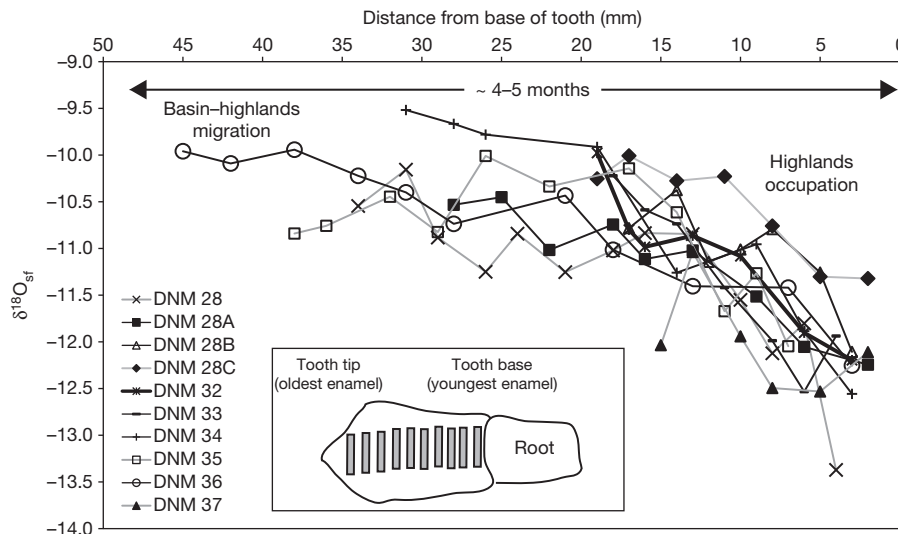


Figure 2 | Oxygen isotope ratios of serial enamel samples compared with position relative to the base of the tooth for ten different teeth from DNM. Teeth form incrementally such that the oldest enamel is at the tip and the youngest enamel is at the base (see inset). Most teeth are characterized by a

gradual decrease in oxygen isotope ratios over time. Tooth DNM 36 is unworn and thus preserves the longest temporal record (approximately 4–5 months), whereas records from other worn teeth are truncated to various degrees.

Supplementary Information for details about estimates of $\delta^{18}\text{O}_{\text{sf}}$. Although $\delta^{18}\text{O}_{\text{sf}}$ inferred for all authigenic carbonates have relatively high and limited $\delta^{18}\text{O}$ values, those inferred for *Camarasaurus* have significantly different means and variances (Supplementary Tables 1–3). Most importantly, they preserve a record of the lowest $\delta^{18}\text{O}_{\text{sf}}$ of all carbonates analysed from the Morrison basin (Fig. 1b).

$\delta^{18}\text{O}_{\text{sf}}$ values inferred from *Camarasaurus* that overlap with those from authigenic carbonates are consistent with these animals spending time in the fluvial and wetland environments of the basin. In contrast, lower $\delta^{18}\text{O}$ of surface water and precipitation ($\delta^{18}\text{O}_{\text{sf}} < \text{approximately } -9\text{‰}$) implied by a large proportion of *Camarasaurus* teeth indicates that they occupied non-basinal settings. Low $\delta^{18}\text{O}_{\text{sf}}$ values result from the preferential rainout of ^{18}O from air masses as they rise, cool and lose water while crossing topographic barriers such as volcanic highlands west of the basin (Fig. 1a). Thus, they indicate that animals from both DNM and Thermopolis were drinking water from these high-elevation regions. Although it is possible that ‘extra-basinal’ high-elevation waters could flow into lakes and rivers located in the basin proper, the fact that lake and wetland carbonates do not have low $\delta^{18}\text{O}$ indicates that such recharge did not have a major influence on $\delta^{18}\text{O}_{\text{sf}}$ in the basin (Fig. 1b). Therefore *Camarasaurus* populations in these areas must have directly occupied high-elevation regions for at least part of the year before returning to the basin where they died. To do so, these animals must have migrated approximately 300 km in each direction based on palaeogeographical reconstructions for the Late Jurassic^{2,9} (Fig. 1a).

Patterns in $\delta^{18}\text{O}$ obtained from single teeth provide evidence that this migration was seasonal in nature. Because vertebrate teeth, including those of dinosaurs¹⁴, form incrementally, sequential sampling along the length of a tooth provides a record of $\delta^{18}\text{O}_{\text{sf}}$ ingested during the time of tooth formation^{15–17}. Intra-tooth variations in $\delta^{18}\text{O}_{\text{sf}}$ inferred from camarasaurid teeth of a single DNM *C. lentus* skull (see Supplementary Information) appear to capture slightly less than half of the sinusoidal cycle that is expected for a single year^{15,16}, thus indicating that these teeth formed over approximately 4–5 months (Fig. 2). The specific pattern in $\delta^{18}\text{O}_{\text{sf}}$ implies that this animal moved out of the basin into highland regions over the period of tooth formation, yet the teeth are found in the basin. Such a situation is possible because tooth enamel does not provide an instantaneous record of ingested $\delta^{18}\text{O}_{\text{sf}}$; rather, there is a temporal lag associated with the turnover of oxygen in the body. This lag is of the order of 2 weeks for small mammals¹⁸, and although the length of time is unknown for

sauropod dinosaurs it cannot have been longer than several weeks to a month, otherwise seasonal variations would be obscured altogether. Thus, over 5–6 months, this individual left the basin for the highlands and then returned to the DNM area.

Assuming that *Camarasaurus* migrated in an effort to obtain the food and water they needed to survive, they would have left the basin during the dry season (presumably summer¹⁹) when plant growth was limited and drought might have been common, and then returned in the wet season (presumably winter¹⁹). The fact that the DNM *C. lentus* died before preserving a record of basin $\delta^{18}\text{O}_{\text{sf}}$ in its tooth enamel suggests that it was recently returned, and that it died during the transition from the dry to wet season. The similarity in bulk $\delta^{18}\text{O}$ among DNM teeth from other individuals suggests that other *Camarasaurus* from DNM exhibited similar behaviour. Without well-constrained intra-tooth data from Thermopolis it is not possible to describe *Camarasaurus* migrational patterns in as much detail in this area. However, the fact that $\delta^{18}\text{O}_{\text{sf}}$ inferred from Thermopolis are generally higher than those from DNM (Fig. 1b) could mean that the Thermopolis teeth captured a different part of the seasonal cycle in $\delta^{18}\text{O}_{\text{sf}}$ and thus might have died during a different time of the year, that Thermopolis teeth grew during different year(s), or that these animals visited a different (possibly lower elevation) part of the western highlands.

Overall, the research presented here provides strong support for the hypothesis that *Camarasaurus* could undertake long seasonal migrations. It does not, however, imply that they must have done so. Ongoing studies of other *Camarasaurus* populations and of other sauropods living in different areas will allow us to determine if migrations were a universal characteristic of these animals, or whether it was a behavioural response to environmental stress. In turn it will be possible to address the role that migration might have played in the evolution of sauropod gigantism.

Received 8 August; accepted 19 September 2011.

Published online 26 October 2011.

- Parrish, J., Peterson, F. & Turner, C. E. Jurassic ‘savannah’ – plant taphonomy and climate of the Morrison Formation (Upper Jurassic, Western USA). *Sedim. Geol.* **167**, 139–164 (2004).
- Turner, C. E. & Peterson, F. Reconstruction of the Upper Jurassic Morrison Formation extinct ecosystem—a synthesis. *Sedim. Geol.* **167**, 309–355 (2004).
- Engelmann, G. F., Chure, D. J. & Fiorillo, A. R. The implications of a dry climate for the paleoecology of the fauna of the Upper Jurassic Morrison Formation. *Sedim. Geol.* **167**, 297–308 (2004).

4. Dodson, P., Behrensmeyer, A. K., Bakker, R. T. & McIntosh, J. S. Taphonomy and paleoecology of the dinosaur beds of the Jurassic Morrison Formation. *Paleobiology* **6**, 208–232 (1980).
5. Gat, J. R. Oxygen and hydrogen isotopes in the hydrologic cycle. *Annu. Rev. Earth Planet. Sci.* **24**, 225–262 (1996).
6. Bowen, G. Isoscapes: spatial pattern in isotopic biogeochemistry. *Annu. Rev. Earth Planet. Sci.* **38**, 161–187 (2010).
7. Kohn, M. J. & Cerling, T. E. Stable isotope compositions of biological apatite. *Rev. Mineral. Geochem.* **48**, 455–488 (2002).
8. Amiot, R. *et al.* Latitudinal temperature gradient during the Cretaceous Upper Campanian–Middle Maastrichtian: $\delta^{18}\text{O}$ record of continental vertebrates. *Earth Planet. Sci. Lett.* **226**, 255–272 (2004).
9. Dunegan, S. & Turner, C. Regional paleohydrologic and paleoclimatic settings of wetland/lacustrine depositional systems in the Morrison Formation (Upper Jurassic), Western Interior, USA. *Sedim. Geol.* **167**, 269–298 (2004).
10. Lovelace, D. An Upper Jurassic Morrison Formation fire-induced debris flow: taphonomy and paleoenvironment of a sauropod (Sauropoda: *Supersaurus vivianae*) locality, east-central Wyoming. *Bull. New Mexico Mus. Nat. Hist.* **36**, 47–56 (2006).
11. Jennings, D. S. & Lovelace, D. M. Driese, S.G. Differentiating paleowetland subenvironments using a multi-disciplinary approach: an example from the Morrison Formation, South Central Wyoming, USA. *Sedim. Geol.* 10.1016/j.sedgeo.2011.03.005 (2011).
12. Fallin, M. J. *Carbon Isotope Stratigraphy of the Morrison and Cloverly Formations and Assessment of Vertical Color Change in the Morrison Formation, Coyote Basin, Wyoming*. Master's thesis, Univ. Colorado, Boulder (2005).
13. Sellwood, B. W. & Valdes, P. J. Jurassic climates. *Proc. Geol. Assoc.* **119**, 5–17 (2008).
14. Erickson, G. M. Incremental lines of von Ebner and the assessment of tooth replacement rates using growth line counts. *Proc. Natl Acad. Sci. USA* **93**, 14623–14627 (1996).
15. Fricke, H. C. & O'Neil, J. R. Inter- and intra-tooth variation in the oxygen isotope composition of mammalian tooth enamel phosphate: implications for palaeoclimatological and palaeobiological research. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **126**, 91–99 (1996).
16. Kohn, M. J., Schoeninger, M. J. & Valley, J. W. Variability in oxygen compositions of herbivore teeth: reflections of seasonality or developmental physiology? *Chem. Geol.* **152**, 97–112 (1998).
17. Stanton-Thomas, K. J. & Carlson, S. J. Microscale $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ isotopic analysis of an ontogenetic series of the hadrosaurid dinosaur *Edmontosaurus*: implications for physiology and ecology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **206**, 257–287 (2004).
18. Podlesak, D. W. *et al.* Turnover of oxygen and hydrogen isotopes in the body water, CO_2 , hair, and enamel of a small mammal. *Geochim. Cosmochim. Acta* **72**, 19–35 (2008).
19. Rees, P. M., Ziegler, A. M. & Valdes, P. J. in *Warm Climates in Earth History* (eds Huber, B. T. *et al.*) 297–318 (Cambridge Univ. Press, 2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hartman, D. Lovelace and D. Chure for providing context and fossil specimens. The research was funded in part by National Science Foundation EAR-0319041 and Colorado College. We thank G. Olack and A. Colman for the phosphate oxygen isotope analyses, supported by National Science Foundation Major Research Instrumentation award EAR 0923831.

Author Contributions J.H. analysed samples and interpreted data from Dinosaur National Monument; M.E.H. analysed samples, interpreted data from Thermopolis and did the statistical analysis. H.C.F. designed the study, aided in data interpretation and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C.F. (hfricke@coloradocollege.edu).

Additive threats from pathogens, climate and land-use change for global amphibian diversity

Christian Hof^{1,2}, Miguel B. Araújo^{1,2,3}, Walter Jetz⁴ & Carsten Rahbek¹

Amphibian population declines far exceed those of other vertebrate groups, with 30% of all species listed as threatened by the International Union for Conservation of Nature^{1–3}. The causes of these declines are a matter of continued research, but probably include climate change, land-use change and spread of the pathogenic fungal disease chytridiomycosis^{4,5}. Here we assess the spatial distribution and interactions of these primary threats in relation to the global distribution of amphibian species. We show that the greatest proportions of species negatively affected by climate change are projected to be found in Africa, parts of northern South America and the Andes. Regions with the highest projected impact of land-use and climate change coincide, but there is little spatial overlap with regions highly threatened by the fungal disease. Overall, the areas harbouring the richest amphibian faunas are disproportionately more affected by one or multiple threat factors than areas with low richness. Amphibian declines are likely to accelerate in the twenty-first century, because multiple drivers of extinction could jeopardize their populations more than previous, monocausal, assessments have suggested.

Amphibians are experiencing population declines in all regions of the world^{2,6}. Causes for this global decline have been identified. Among the highest ranking threats are anthropogenic land-use changes, leading to habitat destruction and fragmentation, and the fatal disease chytridiomycosis, which is caused by the chytrid fungus *Batrachochytrium dendrobatidis*. Other threats include climate change, which may interact with chytridiomycosis, environmental pollution, direct exploitation for the food, medicine and pet trades, increase in ultraviolet-B irradiation due to anthropogenic ozone depletion, and the spread of invasive species^{4,5}.

Many studies have assessed how these threats affect amphibian populations and how they may interact at local and regional scales^{7–10}. Recent assessments have used bioclimatic envelope models to project climate change impacts on amphibian diversity on a continental scale^{11,12}. Attempts have also been made to assess the relative importance of different threats for large groups of species^{13,14}, but not in a spatially explicit way. Several hypotheses have suggested that there may be interactions of chytridiomycosis with land-use change¹⁵ and climate change⁹, yet no final consensus has been reached^{10,16}. Preliminary models of the global geography of chytridiomycosis under climate change have been provided¹⁷, but an integrative, analytical, spatially explicit assessment at a global scale of the spatial interactions of the most severe threats is urgently needed.

Using a nearly complete global data set of 5,527 amphibian species, we elucidate how the spatial interaction at the global scale of three important threats (climate change, chytridiomycosis and land-use change, see Methods) could affect global amphibian diversity between a baseline period (1980) and late in the present century (2080). Threat from future climate change was estimated as the proportion of species locally losing climatic suitability ('climate losers') per area, as given by species-specific bioclimatic models (see Methods). Because the spatial

distribution of species richness varies considerably among the three amphibian orders Anura (frogs and toads, 'frogs' hereafter), Caudata (salamanders and newts, 'salamanders') and Gymnophiona ('caecilians') (Fig. 1a), we conducted climate change analyses separately for each group. Threat from chytridiomycosis was quantified as the future probability of occurrence of *B. dendrobatidis* from a bioclimatic model projection¹⁷. Lastly, estimates of future land-use change (that is, changes from natural to human-encroached land cover) were based on the projections of the Millennium Ecosystem Assessment^{18–20}.

The three major global threats to amphibians exhibit characteristic, yet disparate geographic patterns (Fig. 1b–d). For frogs, regions with a high projected impact of climate change (regions where high proportions of climate losers in 2080 coincide with a high level of species richness) are the northern Andes and parts of the Amazon and the Cerrado in South America, large areas of sub-Saharan tropical Africa, and a small region in South East Asia (Fig. 1b and Supplementary Fig. 1a). In the northern Andes, which harbour the greatest frog diversity world-wide, the proportion of probable climate losers reaches 166 species (73% of the local frog fauna). Globally, the proportion of frogs likely to become climate losers measured as average per grid cell is 54% ($\pm 10\%$ (standard deviation, s.d.)). For salamanders, western North America, northern Central America and southern and south-eastern Europe are the regions projected to be most heavily affected by climate change, as are some areas in northern South America for caecilians (Fig. 1b). In several areas of Central America, up to 21 species (66% of the local salamander fauna) are projected to lose climatic suitability (global mean $\sim 56\% \pm 15\%$ s.d.) (Fig. 1b). The regions with the highest projected probability of occurrence of chytridiomycosis are located in mostly temperate climates as well as mountainous and coastal regions (Fig. 1c). Areas with high projected land-use change are mainly found in tropical Central and South America, tropical Africa and montane parts of central and southern Asia (Fig. 1d).

Geographic coincidence in the intensity of the three types of threat is highly uneven and varies strongly among the three amphibian orders (Figs 1b–d and 2). Within the range occupied by frogs, the spatial overlap between the top 25% affected grid cells for the three threat types is small: 6.1% of these cells (out of a 25% possible) overlap for chytridiomycosis and climate change, 9.1% for climate change and land-use change and 8.6% for chytridiomycosis and land-use change (Figs 2 and 3; see Supplementary Table 1 for sensitivity analyses and spatial null model tests of overlap; see also Methods and Supplementary Methods for further details). The small degree of overlap between areas of highest impact from chytridiomycosis and climate change arises because of the strong association of *B. dendrobatidis* with humid and cool conditions found in temperate regions²¹ and cool tropical high mountains, whereas the negative impacts from climate change are more prominent in the warm and wet tropics (see Fig. 1a). For salamanders, the comparable figures for spatial overlap are 12.2%, 4.6% and 10.0%, respectively (Fig. 3 and Supplementary Table 1). Globally, more than half of the distributional area of each of the three

¹Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen Ø, Denmark. ²Department of Biodiversity and Evolutionary Biology, National Museum of Natural Sciences, CSIC, C/José Gutiérrez Abascal 2, 28006 Madrid, Spain. ³Rui Nabeiro Biodiversity Chair, CIBIO, University of Évora, Largo dos Colegiais, 7000 Évora, Portugal.

⁴Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06520-8106, USA.

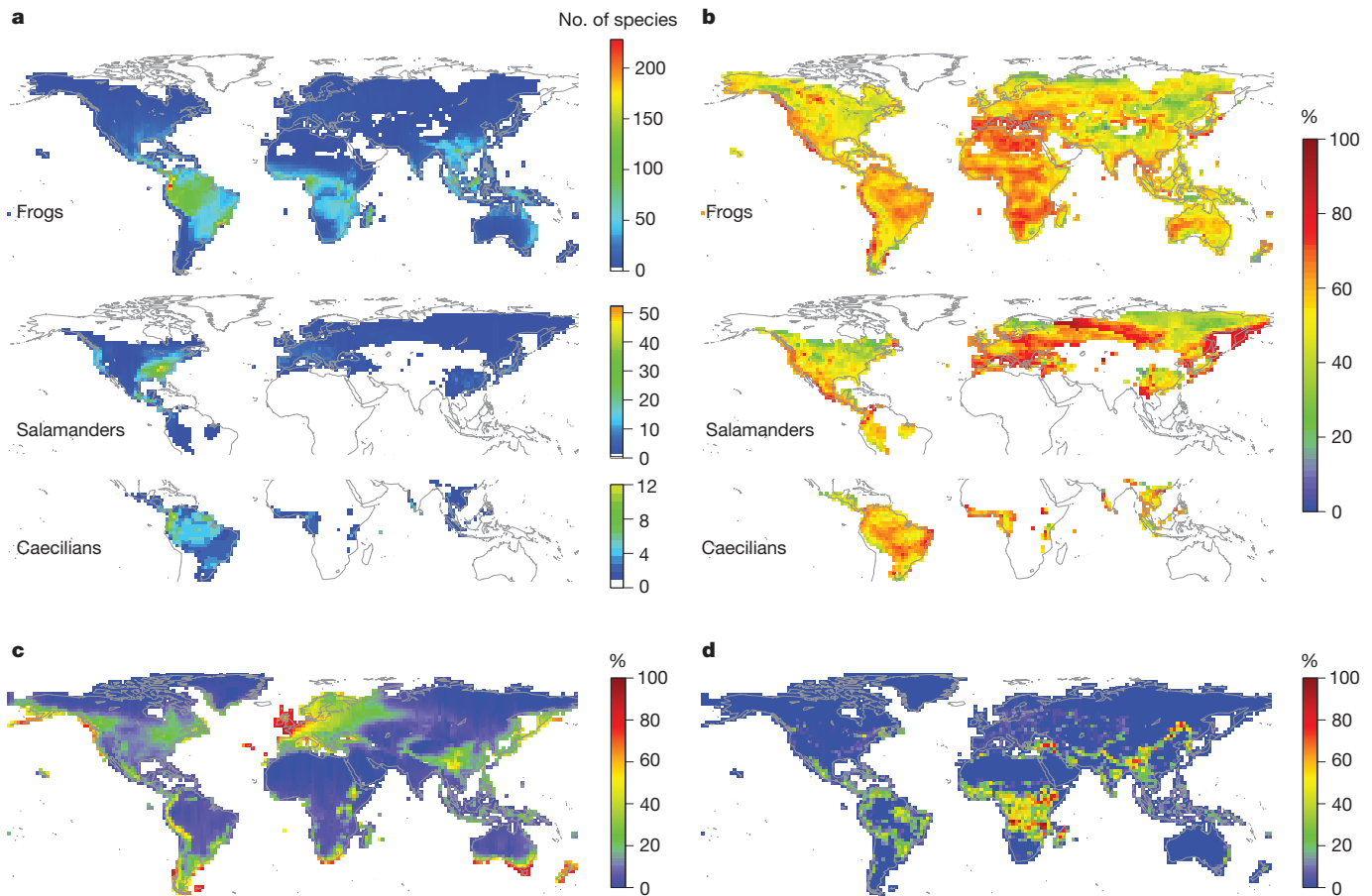


Figure 1 | Current amphibian species richness and the intensity of three factors threatening global amphibian diversity projected for the year 2080. **a**, Spatial variation of species richness (number of species per grid cell) of frogs ($n = 4,875$), salamanders ($n = 508$) and caecilians ($n = 144$). **b**, Intensity of threat from climate change, given as the proportion of species projected to lose climatic suitability in a given area (arithmetic mean across 14 GCMs, 3 emission scenarios and 3 modelling algorithms). **c**, Intensity of threat from

chytridiomycosis, given as the projected probability of occurrence of *B. dendrobatidis* (arithmetic mean across 3 GCMs and 2 emission scenarios, data from ref. 17). **d**, Intensity of threat from land-use change, given as the proportion of a given area projected to be converted from a natural to an anthropogenic state (arithmetic mean across 4 scenarios, data from ref. 19). For further details on the quantification of threat intensities, see Methods. White areas in panels **a** and **b** indicate the absence of a given amphibian order.

amphibian orders is likely to be highly affected by at least one of the three major threat factors. When only regions with the highest species richness are considered, then about two thirds of the areas with the richest frog and salamander faunas (half of the area for caecilians) are projected to become heavily affected by 2080 (Fig. 4).

Several critical assumptions are inevitable when conducting global assessments like ours. Although a careful investigation of some of the methodological uncertainties in our study broadly confirms the consistency of our findings (see Methods, Supplementary Methods and Supplementary Discussion), several limitations are important.

First, much uncertainty exists in ensemble forecasting as illustrated by the results produced by our use of 14 different global climate models (GCMs) and three bioclimatic envelope modelling algorithms (Supplementary Figs 8–12). Although we explicitly examined this variation, results are contingent on the methods used and future modelling might provide further insights. Second, other threats not investigated here may also cause declines in amphibian populations, including pollution, direct exploitation, spread of invasive species, or increased ultraviolet-B irradiation⁵. These additional factors give reason for further concern, especially when they interact with one another⁴. Third,

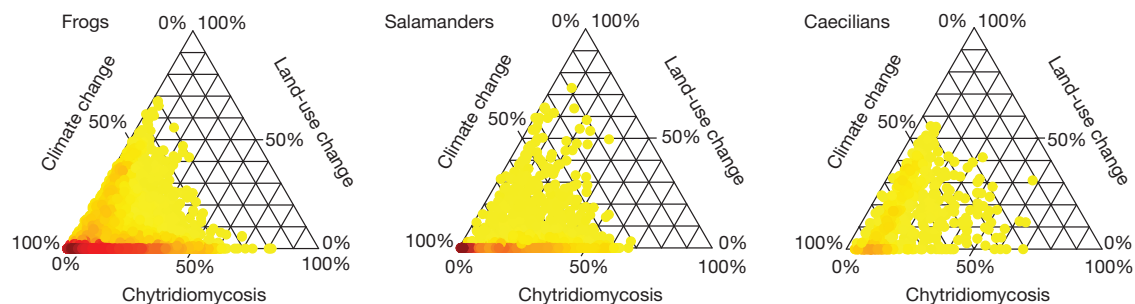


Figure 2 | Relationships among the intensities of the three main factors threatening global amphibian diversity. Each point refers to one grid cell in Fig. 1 and represents the relative intensity of each of the three threats when considered jointly: climate change (proportion of species losing climatic suitability); chytridiomycosis (probability of occurrence of chytridiomycosis);

and land-use change. For details on the quantification of threat intensities, see Fig. 1 and Methods. Values of threat intensity are standardized to vary between 0 and 100%, and then transformed into relative proportions to add up to 100% (division by the sum of the three threat intensities). Darker colours indicate higher point densities.

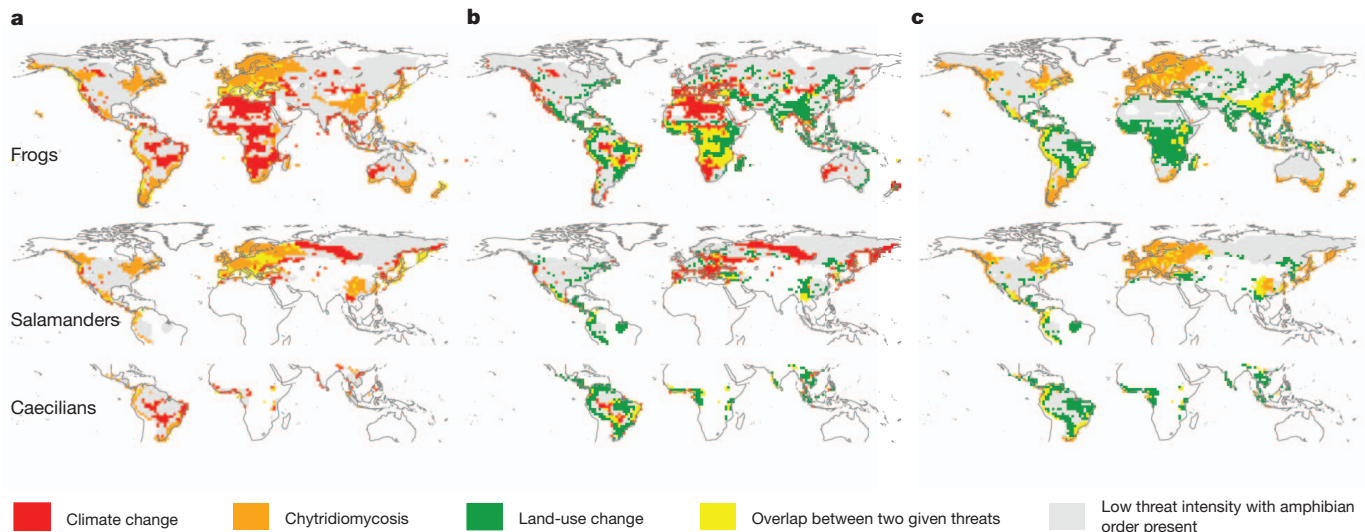


Figure 3 | Spatial distribution and pairwise overlap of the three main factors threatening global amphibian biodiversity, projected for the year 2080.

a, Climate change and chytridiomycosis. **b**, Climate change and land-use change. **c**, Chytridiomycosis and land-use change. Colours indicate areas of particularly high threat intensity. These areas are defined as the 25% of all grid

modelled presence of chytridiomycosis does not always cause local declines or extinctions and interspecific as well as regional differences in amphibian susceptibility to chytridiomycosis exist²². Fourth, the spatial resolution of our analysis is relatively coarse, yet several factors allow

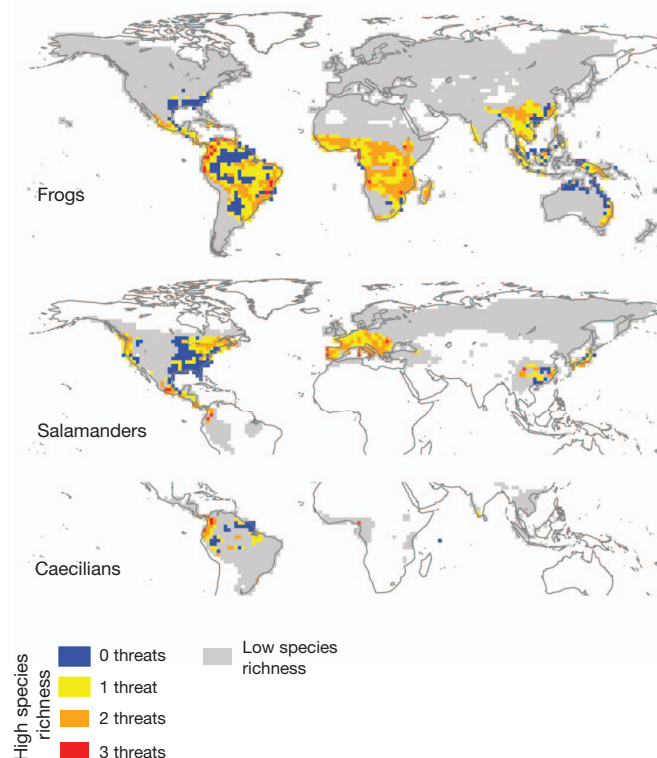


Figure 4 | Spatial overlap between areas with the highest amphibian species richness and the main factors threatening global amphibian diversity, projected for 2080. Areas with the highest species richness are defined as the 25% of all grid cells with the highest number of species. Areas where high levels of species richness coincide with high intensity of 1–3 of the main factors threatening amphibian diversity are coloured in yellow, orange and red, respectively (blue: no coincidence of high richness and high threat intensity). For definition of 25% areas of high threat intensity and further details see Figs 1 and 3.

cells with (1) the highest proportion of species projected to lose climatic suitability, (2) the highest projected probability of occurrence of chytridiomycosis and (3) the highest proportion of land projected to be converted from a natural to an anthropogenic state. For other details see Fig. 1.

population persistence at finer scales (see also discussion in refs 23, 24). In particular, our analyses disregard dispersal, biotic interactions and adaptive potential, and rely on the assumption that coarse-scale projections from bioclimatic envelope models provide a good surrogate for species' climatic requirements at finer scales. However, in some cases, species might persist owing to increased climatic stability in fine-grain refugia²⁵, or to local adaptations of amphibian populations²⁶. Our projections thus provide measurements of the exposure of global amphibian species distributions to key threatening processes but, because the species' responses to these threats are not investigated, projections overestimate the impacts of multiple threats on the persistence of local amphibian populations (see also Supplementary Discussion).

Despite these uncertainties, our results reveal an intriguing pattern of non-overlap between key threatening factors. The implication is that risk assessments focusing on a single threat, such as just climate change or only chytridiomycosis, are probably picturing an optimistic view. As such, they fail to identify the key actions required to curb the ongoing global decline in amphibian diversity. The low coincidence between regions projected to have high prevalence of chytridiomycosis and climate or land-use change emphasizes the potential for silent extinctions away from the regions where the current human footprint is larger. In turn, the higher coincidence between land-use and climate change highlights the existence of potential synergies between the two threatening factors that are often neglected^{18,27,28}. The substantial overlap of threats with many of the world's centres of amphibian richness further underlines the pessimistic long-term perspective for global amphibian diversity^{1,5}. It reinforces the realization that prioritization of conservation efforts needs to be based on knowledge of the spatial distribution both of the different key threats and of biodiversity.

METHODS SUMMARY

To identify the regions with the highest projected impacts of climate change on amphibian species, we fitted three familiar bioclimatic models for each of 5,527 species using an ensemble forecasting framework²⁹. Climate data were extracted from 14 different GCMs (Supplementary Table 2)³⁰ under three emission scenarios. Across 5,041 2×2 degrees latitude–longitude cells, we identified species projected to lose climatic suitability in each grid cell ('climate losers') in 2080 compared with baseline conditions (1980). We then mapped the proportion of climate losers out of the total number of species per grid cell across the world. The probability of occurrence of *B. dendrobatidis* was obtained from a climate-based ensemble modelling projection (data from ref. 17). To estimate land-use change, we used the proportion of a grid cell projected to be actively transformed by humans in the presence of (but

not necessarily driven by) a changing climate (data from ref. 19; procedure following ref. 27).

Regions with highest projected impacts of climate change were identified as the 25% of the grid cells with the highest proportion of losers for each amphibian order. We also used the 25% threshold to identify the areas with the highest species richness, as well as the regions with the highest projected incidence of chytridiomycosis and the highest projected land-use change in 2080. We repeated the analyses using threshold levels of 20% and 10% to assess the consistency of our results. Spatial overlap among grid cells projected to be most affected was assessed by pairwise comparisons of the three threat factors, and also in relation to the distribution of regions with the highest levels of species richness.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 May; accepted 19 October 2011.

Published online 16 November 2011.

- Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
- IUCN. *An Analysis of Amphibians on the 2008 IUCN Red List* (<http://www.iucnredlist.org/initiatives/amphibians>) (2008).
- Wake, D. B. & Vredenburg, V. T. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc. Natl Acad. Sci. USA* **105**, 11466–11473 (2008).
- Blaustein, A. R. & Kiesecker, J. M. Complexity in conservation: lessons from the global decline of amphibian populations. *Ecol. Lett.* **5**, 597–608 (2002).
- Beebee, T. J. C. & Griffiths, R. A. The amphibian decline crisis: a watershed for conservation biology? *Biol. Conserv.* **125**, 271–285 (2005).
- Houlahan, J. E., Findlay, C. S., Schmidt, B. R., Meyer, A. H. & Kuzmin, S. L. Quantitative evidence for global amphibian population declines. *Nature* **404**, 752–755 (2000).
- Becker, C. G., Fonseca, C. R., Haddad, C. F. B., Batista, R. F. & Prado, P. I. Habitat split and the global decline of amphibians. *Science* **318**, 1775–1777 (2007).
- Bosch, J., Carrascal, L. M., Duran, L., Walker, S. & Fisher, M. C. Climate change and outbreaks of amphibian chytridiomycosis in a montane area of Central Spain; is there a link? *Proc. R. Soc. Lond. B* **274**, 253–260 (2007).
- Pounds, J. A. *et al.* Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature* **439**, 161–167 (2006).
- Lips, K. R., Diffendorfer, J., Mendelson, J. R. & Sears, M. W. Riding the wave: reconciling the roles of disease and climate change in amphibian declines. *PLoS Biol.* **6**, e72 (2008).
- Lawler, J. J., Shafer, S. L. & Blaustein, A. R. Projected climate impacts for the amphibians of the Western Hemisphere. *Conserv. Biol.* **24**, 38–50 (2010).
- Araújo, M. B., Thuiller, W. & Pearson, R. G. Climate warming and the decline of amphibians and reptiles in Europe. *J. Biogeogr.* **33**, 1712–1728 (2006).
- Bielby, J., Cooper, N., Cunningham, A. A., Garner, T. W. J. & Purvis, A. Predicting susceptibility to future declines in the world's frogs. *Conservation Letters* **1**, 82–90 (2008).
- Sodhi, N. S. *et al.* Measuring the meltdown: drivers of global amphibian extinction and decline. *PLoS One* **3**, e1636 (2008).
- Becker, C. G. & Zamudio, K. R. Tropical amphibian populations experience higher disease risk in natural habitats. *Proc. Natl Acad. Sci. USA* **108**, 9893–9898 (2011).
- Rohr, J. R., Raffel, T. R., Romansic, J. M., McCallum, H. & Hudson, P. J. Evaluating the links between climate, disease spread, and amphibian declines. *Proc. Natl Acad. Sci. USA* **105**, 17436–17441 (2008).
- Rödger, D., Kielgast, J. & Lötters, S. Future potential distribution of the emerging amphibian chytrid fungus under anthropogenic climate change. *Dis. Aquat. Organ.* **92**, 201–207 (2010).
- van Vuuren, D. P., Sala, O. E. & Pereira, H. M. The future of vascular plant diversity under four global scenarios. *Ecol. Soc.* **11**, 25 (2006).
- Alcamo, J. *et al.* *Global Change Scenarios of the 21st Century, Results from the IMAGE 2.1 Model* 3–96 (Elsevier, 1998).
- Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: Scenarios*. (Island, 2005).
- Daszak, P., Cunningham, A. A. & Hyatt, A. D. Infectious disease and amphibian population declines. *Divers. Distrib.* **9**, 141–150 (2003).
- Kilpatrick, A. M., Briggs, C. J. & Daszak, P. The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends Ecol. Evol.* **25**, 109–118 (2010).
- Hof, C., Rahbek, C. & Araújo, M. B. Phylogenetic signals in the climatic niches of the world's amphibians. *Ecography* **33**, 242–250 (2010).
- Hof, C., Levinsky, I., Araújo, M. B. & Rahbek, R. Rethinking species' ability to cope with rapid climate change. *Glob. Change Biol.* **17**, 2987–2990 (2011).
- Ashcroft, M. B. Identifying refugia from climate change. *J. Biogeogr.* **37**, 1407–1413 (2010).
- Phillimore, A. B., Hadfield, J. D., Jones, O. R. & Smithers, R. J. Differences in spawning date between populations of common frog reveal local adaptation. *Proc. Natl Acad. Sci. USA* **107**, 8292–8297 (2010).
- Jetz, W., Wilcove, D. S. & Dobson, A. P. Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biol.* **5**, e157 (2007).
- Beaumont, L. J. *et al.* Impacts of climate change on the world's most exceptional ecoregions. *Proc. Natl Acad. Sci. USA* **108**, 2306–2311 (2011).
- Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 (2007).
- Meehl, G. A. *et al.* The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**, 1383–1394 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. Rödder, S. Lötters and J. Kielgast for the provision of data and *B. dendrobatidis* modelling results. We thank C. Graham, R. Colwell, N. Sanders, H. H. Bruun and S. Fritz for comments on previous versions of the manuscript. Special thanks to T. Rangel for technical and statistical support. C.H., M.B.A. and C.R. acknowledge the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate; research by M.B.A. was funded by the Portuguese Foundation for Science and Technology (PTDC/AAC-AMB/98163/2008); W.J. acknowledges support from NSF grants DBI 0960550 and DEB 1026764.

Author Contributions C.H., M.B.A. and C.R. designed the study, C.H. performed all analyses, all authors discussed the results. C.H. wrote the paper, with substantial contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.H. (christian.hof@senckenberg.de).

METHODS

Data. We fitted bioclimate envelope models (BEMs) for 5,527 amphibian species from the three amphibian orders Anura (frogs and toads), Caudata (salamanders and newts) and Gymnophiona (caecilians), which for simplicity are henceforth referred to as frogs, salamanders and caecilians. Distribution data were compiled from the Global Amphibian Assessment²¹. Polygons of species' ranges were resampled to a $2^\circ \times 2^\circ$ degree latitude–longitude grid (referred to as the 2° -grid) including 5,041 terrestrial cells (for maps of species richness, see Fig. 1a). This resolution approximates the average of the original resolutions of the climate data sets (see later).

Climatic data were obtained from the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model data set³⁰ of the Fourth Intergovernmental Panel on Climate Change (IPCC) report. Data were derived from 14 coupled Atmosphere–Ocean GCMs and 3 emission scenarios (see Supplementary Table 2 for an overview of the used data sets). Using this series of GCMs we encompass a wide range of equilibrium climate sensitivity (ECS; $2.1\text{--}4.3^\circ\text{C}$; see Supplementary Table 2 for details) and an array of original spatial resolutions, from $1.1^\circ \times 1.1^\circ$ to $3.75^\circ \times 3.75^\circ$ degrees latitude–longitude in the original sets. Outputs for each model were obtained for three SRES³² emission scenarios A1B, A2 and B1, but A2 and B1 scenarios were not available for all of the GCMs (Supplementary Table 2; see ref. 33 for a description of the different scenario storylines). Inclusion of these three scenarios ensures that the models cover a wide range of likely climatic changes.

For each of the GCMs and emission scenarios, five climatic variables were obtained to characterize the baseline period (averaged across a 30-year time period from 1970 to 1999) that was used to calibrate the models. The same variables were then used to make projections into a 30-year time period between 2070 and 2099, subsequently referred to as 2080. The variables used were mean annual rainfall and precipitation seasonality, annual temperature range, minimum temperature and maximum temperature. These variables are known to impose constraints on amphibian physiology and survival³⁴ and are often used to model amphibian species distributions and richness^{12,35}. All climate variables were resampled to the 2° -grid.

By using a global extent approach and given the grain of our analysis, we (1) avoid asserting artificial data quality by inappropriate downscaling of the climatic data, and (2) minimize the problem of false absences in the species distribution data³⁶ as many species in the data set have only been identified in a few localities, with no knowledge about the true occurrence of the species². However, the coarse resolution precludes detailed local assessments of threat interactions and processes; therefore our focus is on documenting coarse spatial patterns.

Projections of the probability of occurrence of *B. dendrobatidis* were obtained from climate-based consensus projections previously described¹⁷. These projections (standardized probability of occurrence given by a consensus of MaxEnt BEMs across three GCMs and two emission scenarios; Supplementary Fig. 3) were resampled to the 2° -grid by weighted averaging. In contrast to the climate change and land-use change projections, for chytridiomycosis we did not use values of the change of the probability of occurrence, as many regions are not infected yet by chytridiomycosis, which makes the absolute value of probability of occurrence a better estimate of future risk of *B. dendrobatidis* infection. For the subsequent analyses, we used a consensus map calculated as the arithmetic mean across all the projections (Fig. 1b). Generally, averaging across different scenarios may be problematic. However, for practical reasons and because separate maps of the variation of the probability of chytridiomycosis occurrence for each combination of used GCM \times scenario did not show strong differences in the spatial pattern (Supplementary Fig. 3), we used the consensus map.

For the projections of potential land-use change, we used data from the Millennium Ecosystem Assessment^{20,37}. The Millennium Ecosystem Assessment uses four scenarios representing a variety of socio-economic and political futures to estimate future changes in the Earth's land-cover²⁷ ('Adapting Mosaic', 'Global Orchestration', 'Order from Strength' and 'TechnoGarden'³²⁰). The Millennium Ecosystem Assessment maps provide information on current and future distributions of 18 different land-cover types at a 0.5° latitude–longitude resolution. For a quantification of potential land-use change we identified grid cells that are projected to change from a natural to an anthropogenic land-cover state (change of any land-cover type to land-cover type 3 'cropland/permanent pasture') and calculated the proportion of area changed for each cell of our 2° -grid for 2080, as a consensus map (arithmetic mean across all four scenarios, Fig. 1c) and separately for each of the four scenarios (Supplementary Fig. 4). As for the projections of chytridiomycosis, for practical reasons and because a separate use of different Millennium Ecosystem Assessment scenarios does virtually not affect the results, we used the consensus map in the subsequent analyses.

Modelling. Three different modelling algorithms, Euclidean distance (ED), Mahalanobis distance (MD) and MaxEnt (MX), were used to run BEMs. These

presence-only algorithms were selected owing to the large number of species with uncertain distributions or very small range sizes to be modelled (see also Supplementary Discussion). The two distance-based methods (ED and MD) measure the similarity of each species' occurrence to the mean (or centre) of the available climatic space. Accordingly species' niches are defined as circular (for ED) or elliptical (for MD) shapes in climatic hyperspace³⁸. BioEnsembles, a computer software which is able to optimize and take advantage of high-speed parallel processing, was used to run the ED and MD models³⁹. MaxEnt version 3.2.4^{40,41}, a machine-learning technique based on the principle of maximum entropy, was used to run the MX models. In MaxEnt, we used a regularization multiplier of 0.5 (a model parameter which allows for adjustment of the degree of model overfitting), because this value represents a balance between being able to fit models for species with very few records while avoiding an unreliable degree of overfitting.

For each of the 5,527 species, we ran each possible modelling combination (3 modelling algorithms \times 14 GCMs \times 3 scenarios \times 2 time periods), which resulted in 1,260,156 models (note that for some GCMs only two scenarios were available, Supplementary Table 2). Standard BEM validation procedures were not applicable in our study (but note that a validation for future scenarios is in any case not possible⁴²). However, we cautiously assessed patterns of variation in model results that may have resulted from different sources of uncertainty, such as species with small numbers of occurrence records, different modelling algorithms, and variation among GCMs that may result from different resolutions and equilibrium climatic sensitivities, as well as different emission scenarios (see Supplementary Discussion, also for discussion on model-inherent assumptions). All analyses were performed separately for frogs, salamanders and caecilians.

Processing of modelling results. We used a no-dispersal scenario as the basic underlying assumption for the further processing of the modelling results for two reasons. First, it is unlikely that amphibians will be able to fully track changes in climatic conditions by shifts of their distributional ranges⁴³, in particular when thinking of the coarse spatial scale of our analyses (see also our discussion of coarse data implications in the Supplementary Discussion). Second, and more importantly, the ranges of many species are extremely small (see Supplementary Fig. 2 for range-size frequency distributions). Because BEM projections can become unreliable for species with few occurrence records^{44–46}, we refrained from projecting a species' range into areas where the species does not currently occur. Furthermore, to avoid uncertainties associated with the choice of thresholds to convert raw model outputs (that is, distance to the optimal centroids or probabilities of occurrence) into binary estimates of presence and absence⁴⁷, we decided not to use any type of thresholding. Instead, following ref. 48, we used the change in climatic suitability that overlapped with species' current ranges and disregarded suitability or probability scores that did not overlap with existing records for the species. The change in climatic suitability was then calculated as the difference of the climatic suitabilities between current and future conditions (standardized MX probability of occurrence or 1 minus the raw distance (standardized between 0 and 1) for ED and MD, respectively; see Supplementary Fig. 5 for an example). This procedure was repeated for each model combination (algorithm \times GCMs \times scenarios) for each species. Despite the standardization of values of suitability change to a range from 0 to 1, the values are not quantitatively comparable across the different modelling algorithms, which is due to general differences in distance-based (ED, MD) and machine-learning (MX) algorithms as well as to software-inherent differences (for histograms and maps of the mean changes of suitability per grid cell, calculated as the means across all species, see Supplementary Figs 6 and 7). Therefore we used a qualitative approach to identify the regions with the strongest projected impacts of climate change on amphibian diversity⁴⁸: for each model combination, we counted the number of species per grid cell that (1) lose climatic suitability ('climate losers': negative change in climatic suitability between current and future conditions), (2) gain climatic suitability ('climate winners': positive change in climatic suitability between current and future conditions) and (3) show no change in climatic suitability between current and future conditions. Note that doing so implies that species will be counted as 'climate loser' or 'climate winner' regardless of the magnitude of those changes, and that species may be identified as losers in one grid cell and as winners in another. However, as we only analyse these counts in categorical and aggregate form (that is, cells with the highest proportion of losers) we expect results to be robust to this simplification and further provide sensitivity analyses (see also Supplementary Discussion).

To identify the regions with the strongest projected impacts of climate change on amphibian diversity, we built consensus maps of the proportion of climate losers and then identified the 25% of all grid cells with the highest proportion of losers (Fig. 3a). Consensus maps were derived by calculating arithmetic means of the proportion of climate losers across all model combinations (algorithm \times GCM \times scenario) for 2080 (Fig. 1a).

Many studies have shown that species distribution modelling results can vastly differ when using different GCMs, emission scenarios and algorithms^{49,50}. To

assess the uncertainties around the consensus, we mapped the proportion of losers separately as arithmetic means (1) across all combinations of GCM \times algorithm per scenario (Supplementary Fig. 8), (2) across all combinations of GCM \times scenario per algorithm (Supplementary Fig. 9A–C), (3) across all combinations of GCM \times scenario per combination of two algorithms (Supplementary Fig. 9D–F), and (4) across all combinations of algorithm \times scenario per GCM (Supplementary Fig. 10). Furthermore, following a novel uncertainty assessment protocol⁵¹, we assessed the proportion of variation explained by different sources of uncertainty (algorithm, GCM, scenario, their interactions, and the residual uncertainty) by variance partitioning (SS proportion in 3-way ANOVAs; Supplementary Fig. 11), and mapped these proportions of uncertainty (see Supplementary Fig. 12, ref. 51 and Supplementary Discussion for details). In addition, we identified the 25% grid cells with the highest proportion of climate losers separately for each model combination and calculated the number of models per grid cell that identified this grid cell as one of the 25% with the highest proportion of losers. These overlap maps were constructed for each possible algorithm combination (ED \times MD \times MX; ED \times MD, ED \times MX, MD \times MX; ED, MD, MX) to assess also the amount of uncertainty that is associated with the use of one, two or three modelling algorithms (Supplementary Fig. 13).

Spatial overlap of different threats. To investigate the spatial overlap of different threats for amphibian diversity, we identified the regions with the highest projected impact for each of the respective threat: 25% of all grid cells with the highest proportion of climate losers (red areas in Supplementary Fig. 1A), 25% of all grid cells with the highest probability of occurrence of *B. dendrobatidis* (orange areas in Supplementary Fig. 1B), and the grid cells with a projected land-use change of at least 25% of the total area (green areas in Supplementary Fig. 1C). All of these calculations were based on the consensus maps derived using the procedures explained above. For statistical analyses of threat overlap, see Supplementary Methods and Supplementary Table 1.

31. IUCN. *Global Amphibian Assessment* (<http://www.iucnredlist.org/initiatives/amphibians/>) (2004).
32. IPCC. *Special Report on Emissions Scenarios, Prepared for the Third Assessment Report* (IPCC, 2000).
33. IPCC. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change: Summary for Policymakers* (IPCC, 2007).
34. Carey, C. & Alexander, M. A. Climate change and amphibian declines: is there a link? *Divers. Distrib.* **9**, 111–121 (2003).
35. Araújo, M. B. *et al.* Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31**, 8–15 (2008).
36. Hurlbert, A. H. & Jetz, W. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl Acad. Sci. USA* **104**, 13384–13389 (2007).
37. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: Synthesis* (Island, 2005).
38. Farber, O. & Kadmon, R. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecol. Modell.* **160**, 115–130 (2003).
39. Bioensembles: software for computer intensive ensemble forecasting of species distributions under climate change v. 1.0 (privately distributed, Madrid, Goiás, Évora, 2009).
40. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* **190**, 231–259 (2006).
41. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175 (2008).
42. Araújo, M. B., Pearson, R. G., Thuiller, W. & Erhard, M. Validation of species–climate impact models under climate change. *Glob. Change Biol.* **11**, 1504–1513 (2005).
43. Smith, M. A. & Green, D. M. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? *Ecography* **28**, 110–128 (2005).
44. Stockwell, D. R. B. & Peterson, A. T. Effects of sample size on accuracy of species distribution models. *Ecol. Modell.* **148**, 1–13 (2002).
45. McPherson, J. M., Jetz, W. & Rogers, D. J. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *J. Appl. Ecol.* **41**, 811–823 (2004).
46. Wisz, M. S. *et al.* Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **14**, 763–773 (2008).
47. Nenzén, H. K. & Araújo, M. B. Choice of threshold alters projections of species range shifts under climate change. *Ecol. Modell.* **222**, 3346–3354 (2011).
48. Araújo, M. B., Alagador, D., Cabeza, M., Nogueira-Bravo, D. & Thuiller, W. Climate change threatens European conservation areas. *Ecol. Lett.* **14**, 484–492 (2011).
49. Araújo, M. B., Whittaker, R. J., Ladle, R. J. & Erhard, M. Reducing uncertainty in projections of extinction risk from climate change. *Glob. Ecol. Biogeogr.* **14**, 529–538 (2005).
50. Pearson, R. G. *et al.* Model-based uncertainty in species range prediction. *J. Biogeogr.* **33**, 1704–1711 (2006).
51. Diniz-Filho, J. A. F. *et al.* Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**, 897–906 (2009).

The *Medicago* genome provides insight into the evolution of rhizobial symbioses

Nevin D. Young^{1*}, Frédéric Debellé^{2,3*}, Giles E. D. Oldroyd^{4*}, Rene Geurts⁵, Steven B. Cannon^{6,7}, Michael K. Udvardi⁸, Vagner A. Benedito⁹, Klaus F. X. Mayer¹⁰, Jérôme Gouzy^{2,3}, Heiko Schoof¹¹, Yves Van de Peer¹², Sebastian Proost¹², Douglas R. Cook¹³, Blake C. Meyers¹⁴, Manuel Spannagl¹⁰, Foo Cheung¹⁵, Stéphane De Mita⁵, Vivek Krishnakumar¹⁵, Heidrun Gundlach¹⁰, Shiguo Zhou¹⁶, Joann Mudge¹⁷, Arvind K. Bharti¹⁷, Jeremy D. Murray^{4,8}, Marina A. Naoumkina⁸, Benjamin Rosen¹³, Kevin A. T. Silverstein¹⁸, Haibao Tang¹⁵, Stéphane Rombauts¹², Patrick X. Zhao⁸, Peng Zhou¹, Valérie Barbe¹⁹, Philippe Bardou^{2,3}, Michael Bechner¹⁶, Arnaud Bellec²⁰, Anne Berger¹⁹, Hélène Bergès²⁰, Shelby Bidwell¹⁵, Ton Bisseling^{5,21}, Nathalie Choisne¹⁹, Arnaud Couloux¹⁹, Roxanne Denny¹, Shweta Deshpande²², Xinbin Dai⁸, Jeff J. Doyle²³, Anne-Marie Dudez^{2,3}, Andrew D. Farmer¹⁷, Stéphanie Fouteau¹⁹, Carolien Franken⁵, Chrystel Gibelin^{2,3}, John Gish¹³, Steven Goldstein¹⁶, Alvaro J. González²⁴, Pamela J. Green¹⁴, Asis Hallab²⁵, Marijke Hartog⁵, Axin Hua²², Sean J. Humphray²⁶, Dong-Hoon Jeong¹⁴, Yi Jing²², Anika Jöcker²⁵, Steve M. Kenton²², Dong-Jin Kim^{13,27}, Kathrin Klee²⁵, Hongshing Lai²², Chunting Lang⁵, Shaoping Lin²², Simone L. Macmil²², Ghislaine Magdelenat¹⁹, Lucy Matthews²⁶, Jamison McCorrison¹⁵, Erin L. Monaghan¹⁵, Jeong-Hwan Mun^{13,28}, Fares Z. Najjar²², Christine Nicholson²⁶, Céline Noirot²⁹, Majesta O'Brien²², Charles R. Paule¹, Julie Poulain¹⁹, Florent Prion^{2,3}, Baifang Qin²², Chunmei Qu²², Ernest F. Retzel¹⁷, Claire Riddle²⁶, Erika Sallet^{2,3}, Sylvie Samain¹⁹, Nicolas Samson^{2,3}, Iryna Sanders²², Olivier Saurat^{2,3}, Claude Scarpelli¹⁹, Thomas Schiex²⁹, Béatrice Segurens¹⁹, Andrew J. Severin⁷, D. Janine Sherrier¹⁴, Ruihua Shi²², Sarah Sims²⁶, Susan R. Singer³⁰, Senjuti Sinharoy⁸, Lieven Sterck¹², Agnès Viollet¹⁹, Bing-Bing Wang¹, Keqin Wang²², Mingyi Wang⁸, Xiaohong Wang¹, Jens Warfsmann²⁵, Jean Weissenbach¹⁹, Doug D. White²², Jim D. White²², Graham B. Wiley²², Patrick Wincker¹⁹, Yanbo Xing²², Limei Yang²², Ziyun Yao²², Fu Ying²², Jixian Zhai¹⁴, Liping Zhou²², Antoine Zuber^{2,3}, Jean Dénarié^{2,3}, Richard A. Dixon⁸, Gregory D. May¹⁷, David C. Schwartz¹⁶, Jane Rogers³¹, Francis Quétier¹⁹, Christopher D. Town¹⁵ & Bruce A. Roe²²

Legumes (Fabaceae or Leguminosae) are unique among cultivated plants for their ability to carry out endosymbiotic nitrogen fixation with rhizobial bacteria, a process that takes place in a specialized structure known as the nodule. Legumes belong to one of the two main groups of eurosids, the Fabidae, which includes most species capable of endosymbiotic nitrogen fixation¹. Legumes comprise several evolutionary lineages derived from a common ancestor 60 million years ago (Myr ago). Papilionoids are the largest clade, dating nearly to the origin of legumes and containing most cultivated species². *Medicago truncatula* is a long-established model for the study of legume biology. Here we describe the draft sequence of the *M. truncatula* euchromatin based on a recently completed BAC assembly supplemented with Illumina shotgun sequence, together capturing ~94% of all *M. truncatula* genes. A whole-genome duplication (WGD) approximately 58 Myr ago had a major role in shaping the *M. truncatula* genome and thereby contributed to the evolution of endosymbiotic nitrogen fixation. Subsequent to the WGD, the *M. truncatula* genome experienced higher levels of rearrangement than two other sequenced legumes, *Glycine max* and *Lotus japonicus*.

M. truncatula is a close relative of alfalfa (*Medicago sativa*), a widely cultivated crop with limited genomics tools and complex autotetraploid genetics. As such, the *M. truncatula* genome sequence provides significant opportunities to expand alfalfa's genomic toolbox.

Optical mapping indicates that the eight pseudomolecules of assembly Mt3.5 span a physical distance of 375 million base pairs (Mb), and fluorescence *in situ* hybridization indicates they extend from pericentromeres almost to telomeric ends (Supplementary Figs 1 and 2). Altogether, Mt3.5 consists of 2,536 bacterial artificial chromosomes (BACs; Supplementary Tables 1 and 2) with 273 physical gaps (including centromeres, Supplementary Table 3) and 101 internal sequencing gaps. The pseudomolecules contain 246 Mb of non-redundant sequence (Supplementary Table 2) located entirely within the optical map (Supplementary Fig. 3). Another 146 unfinished BACs/BAC pools that cannot be placed on the optical map contribute 17.3 Mb. Regions not represented in pseudomolecules or unanchored BACs were captured through assembly of approximately 40× coverage Illumina sequencing, yielding 104.2 Mb of additional unique sequence. Although not directly tested, the Illumina sequence is expected to lie

¹Departments of Plant Pathology and Plant Biology, University of Minnesota, St Paul, Minnesota 55108, USA. ²INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, BP 52627, F-31326 Castanet-Tolosan CEDEX, France. ³CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, BP 52627, F-31326 Castanet-Tolosan CEDEX, France. ⁴Department of Disease and Stress Biology, John Innes Centre, Norwich NR4 7UH, UK. ⁵Laboratory of Molecular Biology, Department of Plant Science, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. ⁶USDA-ARS Corn Insects and Crop Genetics Research Unit, Ames, Iowa 50011, USA. ⁷Department of Agronomy, Iowa State University, Ames, Iowa 50011, USA. ⁸Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, Oklahoma 73401, USA. ⁹Department of Genetics and Developmental Biology, Plant and Soil Science Division, West Virginia University, Morgantown, West Virginia 26506, USA. ¹⁰MIPS/Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, Ingolstädter Landstrasse 1, Neuherberg, Germany. ¹¹University of Bonn, INRES Crop Bioinformatics, Katzenburgweg 2, 53115 Bonn, Germany. ¹²Department of Plant Systems Biology, VIB, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. ¹³Department of Plant Pathology, University of California, Davis, California 95616, USA. ¹⁴Department of Plant & Soil Sciences and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA. ¹⁵J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA. ¹⁶Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, Wisconsin 53706, USA. ¹⁷National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, New Mexico 87505, USA. ¹⁸Masonic Cancer Center, Biostatistics and Bioinformatics Group, University of Minnesota, Minneapolis, Minnesota 55455, USA. ¹⁹Genoscope/Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 Evry CEDEX, France. ²⁰INRA, Centre National de Ressources Génétiques Végétales (CNRGV), BP 52627, F-31326 Castanet-Tolosan CEDEX, France. ²¹College of Science, King Saud University, Post Office Box 2455, Riyadh 11451, Saudi Arabia. ²²Advanced Center for Genome Technology, Department of Chemistry and Biochemistry, Stephenson Research and Technology Center, University of Oklahoma, Norman, Oklahoma 73019, USA. ²³Department of Plant Biology, Cornell University, Ithaca, New York, 14853 USA. ²⁴Department of Computer & Information Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA. ²⁵Max Planck Institute for Plant Breeding Research, Plant Computational Biology, Carl von Linné Weg 10, 50829 Köln, Germany. ²⁶Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²⁷International Institute for Tropical Agriculture, (c/o P.O. Box 30709 Nairobi, Kenya 00100), Ibadan, Nigeria. ²⁸National Institute of Agricultural Biotechnology, Rural Development Administration, 225 Seodun-dong, Gwonseon-gu, Suwon 441-707, South Korea. ²⁹INRA, Unité de Biométrie et d'Intelligence Artificielle (UBIA), UR875, BP 52627, F-31326 Castanet-Tolosan CEDEX, France. ³⁰Department of Biology, Carleton College, Northfield, Minnesota 55057 USA. ³¹The Genome Analysis Centre, Norwich Research Park, Norwich, Norfolk NR4 7UH, UK.

*These authors contributed equally to this work.

predominantly within the boundaries of pseudomolecules (see below). On the basis of expressed sequence tag alignments, the combined data sets capture ~94% of expressed genes, providing a highly informative platform for analysing the euchromatin of *M. truncatula*, although still at the draft stage.

Altogether there are 62,388 gene loci in Mt3.5 (Supplementary Table 4 and Supplementary Fig. 4), with 14,322 gene predictions annotated as transposons. Pseudomolecules and unassigned BACs contain a total of 44,124 gene loci, 177,271 retroelement-related regions and 26,487 DNA transposons, and non-redundant Illumina assemblies contribute an additional 18,264 genes, 75,777 retrotransposon regions and 8,476 DNA transposons (Supplementary Tables 5–9) along with 1,418 organellar insertions (Supplementary Data 1). For pseudomolecules and unassigned BACs, this translates to 16.8 genes, 67.6 retrotransposons and 10.1 DNA transposons per 100 kilobases (kb). Within Illumina sequence assemblies, gene density (17.1 per 100 kb) and retrotransposon density (72.2 per 100 kb) are similar to pseudomolecules and unassigned BACs, whereas DNA transposon density is lower (8.2 per 100 kb). Similarities in gene and transposon densities between BAC and Illumina sequences support the assertion that the Illumina sequence is euchromatic, although the possibility that some Illumina assemblies come from low-copy regions within heterochromatin can not be excluded. Considering only the 47,845 genes with experimental or database support (Supplementary Table 4), the average *M. truncatula* gene is 2,211 bp in length, contains 4.0 exons, and has a coding sequence of 1,001 bp. These values are similar to those observed previously in *Arabidopsis thaliana* (2,174 bp), *Oryza sativa* (3,403 bp) and *Populus trichocarpa* (2,301 bp)^{4–6}.

Recent analyses of plant genomes indicate a shared whole-genome hexaploidy (WGH) preceding the rosid–asterid split at 140–150 Myr ago⁷. Duplication patterns and genomic comparisons strongly suggest an additional WGD approximately 58 Myr ago in the papilionoids^{8,9}. Near the time of this WGD, papilionoids radiated into several clades, the largest of which split quickly into two subclades, the Hologalegina (including *M. truncatula* and *L. japonicus*) and the millettoids (including *G. max* and other phaseoloids) at about 54 Myr ago². We therefore compared *M. truncatula* pseudomolecules with other sequenced plant genomes to learn more about shared synteny and genome duplication history.

There is significant macrosynteny among *M. truncatula*, *L. japonicus* and *G. max* (Fig. 1 and Supplementary Fig. 5a, b). Conserved blocks, sometimes as large as chromosome arms, span most euchromatin in all three genomes. A given *M. truncatula* region is typically syntenic with one other *M. truncatula* region as a result of the approximately 58-Myr-ago WGD, usually in small blocks showing degraded synteny (Fig. 2 and Supplementary Fig. 6). A given *M. truncatula* region is most similar to two *G. max* regions via speciation at about 54 Myr ago and the *Glycine* WGD at <13 Myr ago¹⁰ and less similar to two other *G. max* regions resulting from the ~58-Myr-ago and <13-Myr-ago WGD events. A *M. truncatula* region is likewise most similar to one *L. japonicus* region via speciation at about 50 Myr ago and less similar to a second *L. japonicus* region as a result of the ~58-Myr-ago WGD. Finally, each *M. truncatula* region and its homeologue typically show similarity to three *Vitis vinifera* regions via the pre-rosid WGH. Exceptions to these patterns could be due to gene losses, gains, or rearrangements specific to the *M. truncatula* lineage, resulting in synteny being more evident between *M. truncatula* and other genomes than in self-comparisons. Indeed, self-comparisons within *M. truncatula* reveal few remnants of the legume-specific WGD (Fig. 2 and Supplementary Fig. 6). Whereas this seems paradoxical, it is probably explained by extensive gene fractionation between WGD-derived homeologues in *M. truncatula*. In Fig. 3, two short regions on Mt1 and Mt3 resulting from the ~58-Myr-ago WGD are displayed beside microsyntenic regions of *G. max* and *V. vinifera*. As expected, many genes are microsyntenic between *M. truncatula* and *G. max* (ranging from 7/19 between Mt3 and Gm14 to 10/20 between Mt1 and Gm17).

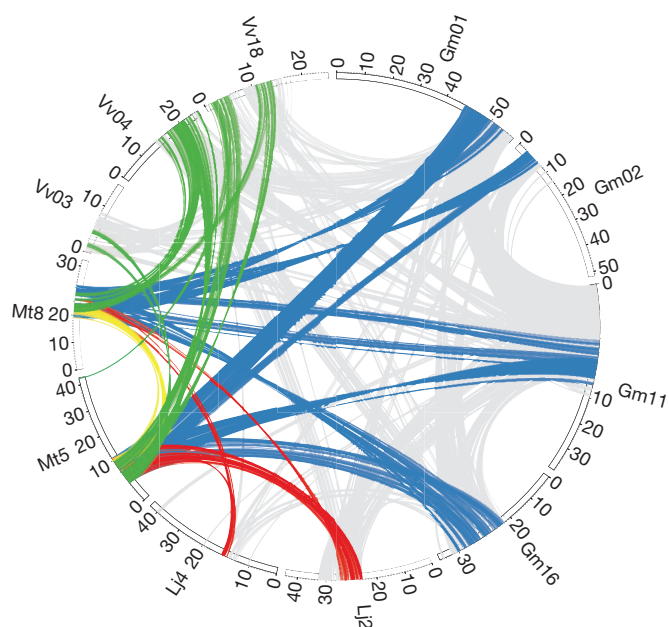


Figure 1 | Circos diagram illustrating syntenic relationships between *Medicago*, *Glycine*, *Lotus* and *Vitis*. Homologous gene pairs were identified for all pairwise comparisons between *M. truncatula*, *G. max*, *L. japonicus* and *V. vinifera* genomes. Syntenic regions associated with the ancestral WGD events were identified by visually inspection of corresponding dot-plots. The large Mt5–Mt8 synteny block (yellow) was found to have two syntenic regions in *L. japonicus* (red), four syntenic regions in *G. max* (blue) and three in *V. vinifera* (green).

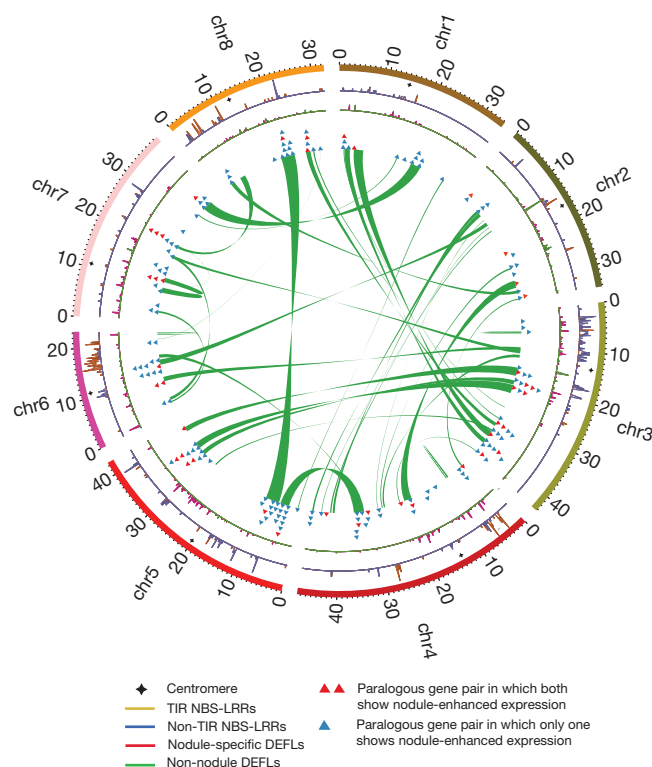


Figure 2 | Circos diagram illustrating the *Medicago* WGD and selected gene families. The 963 WGD-derived paralogous gene pairs were examined for overlap with the nodule-enhanced gene list (Supplementary Data 2). Resulting gene pairs were joined and plotted as either blue triangles (only one of the duplicates is nodule-enhanced) or red (both nodule enhanced). Gene densities of NBS-LRRs, NCRs and other defensin-like proteins are plotted against chromosome position. Density was calculated using a sliding window (100-kb window with 50-kb steps).

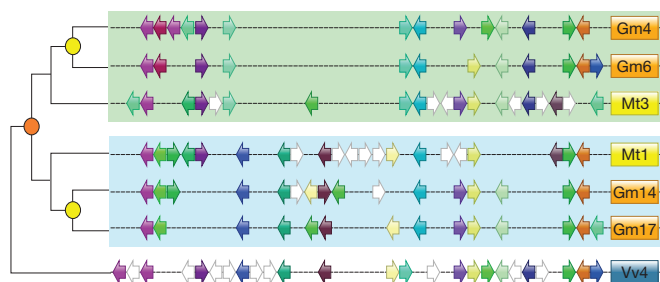


Figure 3 | Microsynteny comparison between *Medicago* homeologues and corresponding regions of *Glycine* and *Vitis*. Microsyntentic genome segments are centred around Medtr3g104510/Medtr1g015890 (Supplementary Table 10), a duplicated region derived from the ~58-Myr-old WGD event noted in orange. The <13-Myr-old *G. max*-specific WGD is coloured yellow. Orthologous/paralogous gene pairs are indicated through use of a common colour. White arrows represent genes with no syntentic homologue(s) in this genome region. Some of these genes may actually have a syntentic sequence in soybean but no corresponding model reported in the current annotation (<http://www.phytozome.net/soybean>).

Between the two *M. truncatula* homeologues, however, only 6 out of 33 genes (or collapsed gene families) are microsyntentic, with a homeologue missing from one or the other duplicate (Supplementary Table 10). Apparently, there have been many more changes, large and small, in *M. truncatula* than in *G. max* since the legume WGD. This is borne out by the fact that synteny blocks in *M. truncatula* are one-third the length of those remaining from the papilionoid WGD in *G. max* (524 kb against 1,503 kb) with the average number of homologous gene pairs per block correspondingly lower (12.4 against 31.0).

The *M. truncatula* genome also has undergone high rates of local gene duplication. The ratio of related genes within local clusters compared to all genes in families is 0.339 in *M. truncatula*, 3.1-fold higher than in *G. max* and 1.6-fold higher than in *A. thaliana* or *P. trichocarpa*. ('Local clusters' are defined as genes in a family all within 100 gene models of one another.) The excess of local gene duplications in *M. truncatula* is observed genome-wide and affects many families. There are 2.63 times as many gene families with local duplications in *M. truncatula* compared with *G. max* (2,980 against 1,131), an excess that also is seen in detailed comparisons of syntentic regions in *M. truncatula* and *G. max*. We examined 16.3 Mb of Mt05 showing synteny to two large regions of Gm01 plus homeologous blocks on Gm02, Gm09 and Gm11. In these regions, 25.8% of *M. truncatula* genes are locally duplicated compared with just 8.0% in *G. max*. Local gene duplications and losses have contributed both to synteny disruptions (Fig. 3 and Supplementary Fig. 7) and to high gene count (62,388) in *M. truncatula*—a value nearly as high as the 65,781 total gene models in *G. max* despite its additional (<13 Myr ago) WGD. Local gene duplications are evident in certain gene families, such as F-box genes, which have undergone pronounced expansions (Supplementary Fig. 8 and Supplementary Table 11). *M. truncatula* also has experienced higher rates of base substitution compared to other plant genomes (Supplementary Fig. 9). Assuming 58 Myr ago as the date of the legume WGD, then the rate of synonymous substitutions per site per year in *M. truncatula* is 1.08×10^{-8} , 1.8 times faster than estimates in other vascular plants¹¹. Higher rates of mutation and greater levels of rearrangement in *M. truncatula* following the papilionoid duplication may have been driven by factors including short generation times, high selfing rates or small effective population sizes, although these characteristics are not unique to *M. truncatula*.

Legumes and actinorhizal species are capable of forming a specialized organ, the root nodule, a highly differentiated structure hosting nitrogen-fixing symbionts. Phylogenetic studies suggest that nodulation may have evolved multiple times in the Fabidae, but the observation that all nodulating species are contained within this single clade indicates

that a predisposition to nodulate evolved in their common ancestor¹². It is unknown whether nodulation with rhizobia preceded the divergence of the three legume subfamilies or evolved on multiple occasions¹³. Nevertheless, rhizobial nodulation and the 58-Myr-old WGD are features common to most papilionoid legumes and both occurred early in the emergence of the group². Given that WGDs generate genetic redundancy that potentially facilitates the emergence of novel gene functions without compromising existing ones¹⁴, we examined the *M. truncatula* genome to ask whether the 58-Myr-old WGD might have had a role in the evolution of rhizobial nodulation in *M. truncatula* and its relatives.

Nod factors are bacterial signalling molecules that initiate nodulation. Previous studies have shown that several of the plant components involved in the response to Nod factors also function in mycorrhizal signalling¹⁵. However, some Nod factor receptors and transcription factors have distinctly nodulation-specific functions. Among these nodulation-specific components, we found that the Nod factor receptor, *NFP*, and the transcription factor, *ERN1*, each have paralogues, *LYR1* and *ERN2* respectively, that trace back to the papilionoid WGD based on genome location and synonymous substitution rate values (Supplementary Fig. 10 and Supplementary Data 2). Both sets of gene pairs also show contrasting expression patterns and functional specialization. *NFP* and *ERN1* are expressed predominantly in the nodule and are known to function in nodulation^{16,17}, whereas *LYR1* and *ERN2* are highly expressed during mycorrhizal colonization (Supplementary Fig. 11). These observations indicate that two important nodulation-specific signalling components in *M. truncatula* might have evolved from more ancient genes originally functioning in mycorrhizal signalling and then duplicated by the 58-Myr-old WGD. In the case of *M. truncatula* *NFP*/*LYR1*, this conclusion is supported by the observation that the apparent orthologue of *NFP* in the nodulating non-legume *Parasponia andersonii* functions in both nodule and mycorrhizal signalling¹⁸. Thus, the 58-Myr-old WGD seems to have led to sub-functionalization of an ancestral gene participating in both interactions, resulting in two homeologous genes that each performs just one of the original functions.

To assess further the contribution of the WGD to *M. truncatula* nodulation, we analysed expression of paralogous gene pairs using RNA-seq data from six different organs (Supplementary Methods 5.1). A total of 963 WGD-derived gene pairs were found (Supplementary Data 2) with 618 pairs (1,046 genes) having RNA-seq data for one or both homeologue. We then determined the number of genes showing organ-enhanced expression (defined as genes with expression level in a single organ at least twice the level in any other) within the pseudomolecule and the WGD-derived gene sets (Supplementary Table 12). In both cases, different organs contained markedly different numbers of genes with enhanced expression (χ^2 with 5 degrees of freedom, $P = 10^{-272}$); however, the rank order among the organs was identical. Roots had the largest number of genes with enhanced expression followed by flower, nodule, leaf, seed/pod and bud. Among gene pairs with nodule-enhanced expression, both paralogues were nodule-enhanced in eight pairs, whereas just a single paralogue was nodule-enhanced in the other 43 pairs. This is consistent with nodulation pre-dating the WGD and further sub- and neo-functionalization emerging afterwards. We went on to examine transcription factors because they can act as regulators of plant growth and development. A total of 3,692 putative TF genes were discovered (Supplementary Data 3), representing 5.9% of all *M. truncatula* gene models (Supplementary Table 13). Of the 1,513 TF genes on pseudomolecules with RNA-seq data, 142 genes (9.4%) derived from the 58-Myr-old WGD (Supplementary Fig. 12 and Supplementary Data 4), consistent with previous observations indicating greater retention of transcription factors following polyploidy¹⁹. Nodule-enhanced expression was significantly higher among transcription factors (92 out of 1,513 or 6.1%) than among all pseudomolecule genes (1,111 out of 23,478 or 4.7%) (χ^2 with 1 degree of freedom, $P = 0.024$) (Supplementary Table 12).

Nodule-enhanced expression was even higher in WGD-derived transcription factors (11 out of 142 or 7.7%), although this enrichment did not reach statistical significance ($P = 0.113$). As expected, *ERN1* is found within this group of WGD-retained, nodule-enhanced transcription factors.

These results show that many paralogous genes retained from the 58-Myr-ago WGD, especially signalling components and regulators, have undergone sub- or neo-functionalization, including several with specialized roles in nodulation. Nevertheless, separate phylogenetic analyses (Supplementary Methods 5.5) indicate that some nodule-related genes derive from the more ancient pre-rosid WGH, with their nodule-related functions pre-dating the 58-Myr-ago WGD (Supplementary Data 5). Taken together, these results are consistent with a model where the capacity for primitive interaction with new symbionts derived from existing mycorrhizal machinery involving genes recruited from the pre-rosid WGH. This capacity would have arisen early in the Fabidae clade and led to the appearance of nodulation in multiple lineages^{13,20}. Later, the 58-Myr-ago WGD would have resulted in additional genes, including *NFP*, *ERN1* and the transcription factors described above, that went on to become specialized for nodule-related functions in the Papilionoideae.

Medicago contains additional amplified gene families, many nodulation-related and found in tandem clusters. *M. truncatula* has nine symbiotic leghaemoglobins, more than twice the number in *L. japonicus* or *G. max* (Supplementary Fig. 13). Five of these genes are located in a tight cluster on Mt5. The *M. truncatula* genome contains 593 nodule cysteine-rich peptides (NCRs) (Supplementary Data 6), a gene family restricted to *M. truncatula* and its relatives²¹. NCRs are noteworthy because they include members essential for terminal differentiation of rhizobia²². NCRs are tightly clustered within the *M. truncatula* genome (Fig. 2), with 75% found in clusters of up to 11 members. The *M. truncatula* genome also has 764 nucleotide-binding site and leucine-rich repeat (NBS-LRR) genes (Supplementary Data 7), more than other plant genomes that have been sequenced so far^{23–25}, many with nodule-specific expression (Supplementary Fig. 14). Almost 90% of NBS-LRRs occur in clusters and genome regions showing limited macrosynteny to other species, such as Mt3 and Mt6, are locations of large NBS-LRR superclusters (Fig. 2 and Supplementary Tables 14 and 15). Finally, *M. truncatula* secretes flavonoid signalling molecules to induce the *nod* genes of *Sinorhizobium meliloti*²⁶. In *M. truncatula*, the corresponding biosynthetic pathway has expanded markedly, with 28 *M. truncatula* chalcone synthase genes in clusters of up to seven members compared to just four chalcone synthases in *A. thaliana*²⁷ (Supplementary Data 8). *M. truncatula* has ten chalcone reductases compared to none in *A. thaliana*²⁸ and *M. truncatula* has 11 chalcone isomerase genes, including one cluster of seven members, compared to just one representative in *A. thaliana*²⁹ (Supplementary Figs 15 and 16).

Analysis of the *M. truncatula* genome supports earlier studies indicating that the dramatic radiation of the legume family (at least the papilionoid subfamily) is partly attributed to the 58-Myr-ago WGD³⁰. Our results indicate that the WGD early in papilionoid evolution allowed the emergence of critical components in Nod factor signalling and contributed to the complexity of rhizobial nodulation observed in this clade. As such, the WGD seems to have had a crucial role in the success of papilionoid legumes, enhancing their utility to humans.

METHODS SUMMARY

DNA sequencing. Six A17 BAC and one fosmid library were used to create Mt3.5 (Supplementary Table 1). Most were processed by Sanger paired-end sequencing of 3–6-kb shotgun libraries. Sequences were downloaded in February/March 2009 with scaffolding performed by aligning all BAC and fosmid ends against contigs and then anchored and ordered primarily by optical mapping. Separately, 25 billion base pairs (Gb) of Illumina sequence was generated using short (375 nt) inserts plus 2.1 Gb from a 5 kb mate-pair library, then assembled using CLCbio (<http://www.clcbio.com>) and Soap (<http://soap.genomics.org.cn/>).

RNA sequencing. Five tissues were used for RNA-seq analysis with ~10 million Illumina 36-bp reads per library (Supplementary Table 12). Three tissues were used for small RNA analysis with ~3 million reads per Illumina library (Supplementary Figs 17–18, Supplementary Table 16 and Supplementary Data 9).

Received 13 June; accepted 13 October 2011.

Published online 16 November 2011.

- Wang, H. *et al.* Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl Acad. Sci. USA* **106**, 3853–3858 (2009).
- Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol.* **54**, 575–594 (2005).
- Kulikova, O. *et al.* Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J.* **27**, 49–58 (2001).
- The Arabidopsis Genome Initiative. I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
- International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
- Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
- Tang, H. *et al.* Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).
- Pfeil, B. E., Schlueter, J. A., Shoemaker, R. C. & Doyle, J. J. Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* **54**, 441–454 (2005).
- Cannon, S. B. *et al.* Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS ONE* **5**, e11630 (2010).
- Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
- Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
- Soltis, D. E. *et al.* Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl Acad. Sci. USA* **92**, 2647–2651 (1995).
- Doyle, J. J. & Luckow, M. A. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol.* **131**, 900–910 (2003).
- Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**, 805–814 (2006).
- Oldroyd, G. E. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59**, 519–546 (2008).
- Arrighi, J. F. *et al.* The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. *Plant Physiol.* **142**, 265–279 (2006).
- Middleton, P. H. *et al.* An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *Plant Cell* **19**, 1221–1234 (2007).
- Op den Camp, R. *et al.* LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**, 909–912 (2011).
- Thomas, B. C., Pedersen, B. & Freeling, M. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* **16**, 934–946 (2006).
- Kistner, C. & Parniske, M. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518 (2002).
- Kato, T. *et al.* Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Mol. Plant Microbe Interact.* **15**, 129–137 (2002).
- Van de Velde, W. *et al.* Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**, 1122–1126 (2010).
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R. W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834 (2003).
- Yang, S., Zhang, X., Yue, J. X., Tian, D. & Chen, J. Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* **280**, 187–198 (2008).
- Zhou, T. *et al.* Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol. Genet. Genomics* **271**, 402–415 (2004).
- Peters, N. K., Frost, J. W. & Long, S. R. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* **233**, 977–980 (1986).
- Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* **126**, 485–493 (2001).
- Hegnauer, R. Relevance of seed polysaccharides and flavonoids for the classification of the leguminosae: a chemotaxonomic approach. *Phytochemistry* **34**, 3–16 (1993).
- Shirley, B. W. *et al.* Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659–671 (1995).
- Singer, S. R. *et al.* Venturing beyond beans and peas: what can we learn from *Chamaecrista*? *Plant Physiol.* **151**, 1041–1047 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding support to N.D.Y., C.D.T. and B.A.R. from The Noble Foundation and NSF-PGRP 0321460, 0604966; to N.D.Y., J.M. and G.D.M. from NSF-PGRP 0820005; to C.D.T. from NSF-PGRP 0821966; to F.D., G.E.D.O., R.G., K.F.X.M., T.B., J. Denarié, F.Q. and J.R. from FP6 EU project GLIP/Grain Legumes

FOOD-CT-2004-506223; to G.E.D.O. and J.R. from BBSRC BBS/B/11524; to F.D. and F.Q. from ANR project SEQMEDIC 2006-01122; to R.G. from the Dutch Science Organization VIDI 864.06.007, ERA-PG FP-06.038A; to Y.V.d.P. from the Belgian Federal Science Policy Office IUAP P6/25, Fund for Scientific Research Flanders, Institute for the Promotion of Innovation by Science and Technology in Flanders and Ghent University (MRP N2N); to D.R.C. from NSF IOS-0531408, IOS-0605251; to D.J.S., B.C.M. and P.J.G. from USDA CSREES 2006-03567; to J. Gouzy from 'Laboratoire d'Excellence' (LABEX) TULIP (ANR-10-LABX-41). We also acknowledge technical support from the University of Minnesota Supercomputer Institute and thank Y. W. Nam for a BamHI BAC library used by Genoscope, S. Park and M. Accerbi for RNA isolation, T. Paape for statistical consulting, and M. Harrison for supplying *myc* infected and control root tissues used to make small RNA libraries.

Author Contributions Planning, coordination and writing: N.D.Y., J. Doyle, F.Q., J. Weissenbach, P.W., K.F.X.M., C.D.T., G.E.D.O., G.D.M., J. Mudge, E.F.R., R.A.D., M.K.U., F.D., J. Denarié, D.R.C., P.J.G., B.C.M., D.J.S., C.R.P., B.A.R., D.C.S., S.B.C., Y.V.d.P., R.G., T.B., J.R., S.R.S.; BAC libraries: B.S., A. Bellec, H.B., J. Gish, D.-J.K.; Mapping and assembly: V.B., N.C., S.F., G.M., S. Samain, E.L.M., F.P., N.S., O.S., A.Z., C.G., J.-H. Mun, R.D., M.B., S.Z., C.L., M.H., C.F., C. Nicholson, C.R.; sequencing: A. Berger, J.P., A.V., D.-H.J., S.D., Y.J., H.L.,

S.L.M., F.Z.N., B.Q., C.Q., M.O., I.S., R.S., K.W., D.D.W., G.B.W., Y.X., L.Y., Z.Y., F.Y., L.Z., S.J.H., L.M., S. Sims; annotation and bioinformatics: A.C., C.S., H.G., M. Spannagl, C. Noirot, T.S., A.J.S., S.B., F.C., V.K., J. McCarrison, H.T., A. Hallab, A.J., K.K., J. Warfsmann, A.K.B., A.D.F., V.A.B., J.D.M., M.A.N., S. Sinharoy, P.X.Z., P.B., A.-M.D., J. Gouzy, E.S., H.S., B.R., A.J.G., J.Z., B.-B.W., X.W., P.Z., K.A.T.S., A. Hua, S.M.K., S.L., J.D.W., S.G., S.P., S.R., L.S., S.D.M., M.W.

Author Information *Medicago truncatula* pseudomolecules are found at DDBJ/EMBL/GenBank as accession numbers CM001217–CM001224 and unanchored BACs as GL982851–GL982996. Illumina genome sequences are in the Short Read Archive under SRS150378, RNA-seq sequences under SRP008485, and small RNA sequences in GEO under GSM769273, GSM769274 and GSM769276. Pseudomolecule annotation and Illumina assemblies are available at ftp://ftp.jcvi.org/pub/data/m_truncatula/Mt3.5/. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike license and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D.Y. (neviny@umn.edu).

Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour

Howard C. Chang¹, Jennifer Paek¹ & Dennis H. Kim¹

Heritable variation in behavioural traits generally has a complex genetic basis¹, and thus naturally occurring polymorphisms that influence behaviour have been defined only in rare instances^{2,3}. The isolation of wild strains of *Caenorhabditis elegans* has facilitated the study of natural genetic variation in this species⁴ and provided insights into its diverse microbial ecology⁵. *C. elegans* responds to bacterial infection with conserved innate immune responses^{6–8} and, although lacking the immunological memory of vertebrate adaptive immunity, shows an aversive learning response to pathogenic bacteria⁹. Here, we report the molecular characterization of naturally occurring coding polymorphisms in a *C. elegans* gene encoding a conserved HECT domain-containing E3 ubiquitin ligase, HECW-1. We show that two distinct polymorphisms in neighbouring residues of HECW-1 each affect *C. elegans* behavioural avoidance of a lawn of *Pseudomonas aeruginosa*. Neuron-specific rescue and ablation experiments and genetic interaction analysis indicate that HECW-1 functions in a pair of sensory neurons to inhibit *P. aeruginosa* lawn avoidance behaviour through inhibition of the neuropeptide receptor NPR-1 (ref. 10), which we have previously shown promotes *P. aeruginosa* lawn avoidance behaviour¹¹. Our data establish a molecular basis for natural variation in a *C. elegans* behaviour that may undergo adaptive changes in response to microbial pathogens.

C. elegans is initially attracted to pathogenic *P. aeruginosa* PA14, but within a few hours develops an avoidance response to *P. aeruginosa* (Supplementary Fig. 1). *P. aeruginosa* lawn avoidance behaviour confers increased survival in the presence of a lawn of pathogenic bacteria^{11–14}. We followed the kinetics of *P. aeruginosa* lawn avoidance behaviour of the laboratory *C. elegans* wild-type strain N2 after transfer from *Escherichia coli* OP50 (Fig. 1a). We observed that over 50% of the N2 population vacated the lawn of *P. aeruginosa* by 8 h, and by 24 h, over 90% of the animals were found outside of the lawn of *P. aeruginosa* (Fig. 1b and Supplementary Fig. 1).

Characterization of the *npr-1* gene has shown that the 215V allele of *npr-1* found in the N2 strain, which has likely been derived in the laboratory^{15,16}, has increased activity compared with the ancestral 215F allele of *npr-1*, modifying several behavioural phenotypes including aggregation, aerotaxis and locomotion specifically in the presence of bacterial food^{10,17–19}. In addition, we have shown previously that the 215V allele of *npr-1* confers N2 with enhanced behavioural avoidance of a lawn of *P. aeruginosa* compared with wild isolates such as CB4856 that carry the 215F allele of *npr-1* (ref. 11). Consistent with these prior observations, we observed that the wild isolates CB4856 and RC301 that carry the 215F allele of *npr-1* exhibited markedly delayed avoidance times from a lawn of *P. aeruginosa* compared with the *P. aeruginosa* avoidance times observed for N2 (Fig. 1b and Supplementary Fig. 1).

We observed that *P. aeruginosa* lawn avoidance behaviour by the CB4856 strain was also diminished relative to RC301, with approximately 75% of CB4856 animals remaining inside the lawn of *P. aeruginosa* after 24 h, compared with 25% lawn occupancy by RC301 animals after 24 h (Fig. 1b and Supplementary Fig. 1). We observed that the DA650 strain¹⁰, which was generated by backcrossing

the 215F *npr-1*(g320) allele of the RC301 strain into the N2 genetic background, showed kinetics of *P. aeruginosa* lawn avoidance comparable to RC301 (Fig. 1b and Supplementary Fig. 1). These observations confirmed that the *npr-1* polymorphism is a major determinant of behavioural avoidance of the *P. aeruginosa* lawn, but also revealed that additional genetic differences among wild isolates of *C. elegans* influence the behavioural avoidance of pathogenic bacteria.

We sought to identify these additional genetic loci that modulate the behavioural avoidance of pathogenic bacteria. We found a substantial contribution from a single locus in determining the different pathogen avoidance behaviours observed in CB4856 and DA650 (Supplementary Fig. 2). We performed positional mapping and transgenic rescue to narrow the genomic interval containing causative genetic differences to an 8-kilobase fragment (genetic analysis and positional mapping detailed in Supplementary Fig. 2). This 8 kb genomic interval included a single gene, F45H7.6 (Fig. 1c), which encodes an E3 ubiquitin ligase with two WW domains and a HECT domain (Fig. 1d), with strong homology to the human HECT, C2, and WW-domain-containing E3 ubiquitin ligase 1 (HECW1, also known as NEDL1)²⁰. Thus, we named the *C. elegans* F45H7.6 gene *hecw-1*. We identified two polymorphisms in *hecw-1* (Fig. 1c), including one coding polymorphism in exon 8 of *hecw-1* that results in a Q325P change in HECW-1 in strain CB4856 (Fig. 1d). Consistent with our mapping data implicating *hecw-1* as a determinant of *P. aeruginosa* lawn avoidance behaviour, we found that a deletion allele of *hecw-1*, *ok1347* (Fig. 1c), showed enhanced *P. aeruginosa* lawn avoidance behaviour (Supplementary Fig. 3).

We sequenced the region of the *hecw-1* gene region encompassing amino acid 325 from a total of 162 *C. elegans* strains (Supplementary Table 1), and we found that only CB4856 carries the *hecw-1* 325P polymorphism. Unexpectedly, we identified a second polymorphism, which results in a Y322C substitution in HECW-1 (Fig. 1d) in 11 wild strains, including AB2 and CB3198. Comparison of the *C. elegans* amino acid sequence flanking and including the polymorphisms with the sequences from other *Caenorhabditis* species and mammalian *hecw-1* orthologues indicated that these polymorphisms lie in a conserved, but relatively rapidly evolving region (Supplementary Fig. 4a).

Whereas the phenotypic effect of the *hecw-1* 325P polymorphism in CB4856 was suggested by mapping and rescue experiments, the *hecw-1* 322C polymorphism was detected by sequencing. To determine the effect of Y322C substitution on pathogen avoidance behaviour, we crossed the *hecw-1* 322C polymorphism in AB2 and CB3198 strains into the CB4856 background (thus substituting the *hecw-1* 322C 325Q alleles for the *hecw-1* 322Y 325P alleles of CB4856). We also crossed the *hecw-1* 322Y 325Q allele of N2 into the CB4856 background. After six backcrosses to the CB4856 strain, we observed that each of the two strains carrying the *hecw-1* 322C 325Q alleles showed *P. aeruginosa* lawn avoidance behaviour that was equivalent to that observed for the CB4856 strain (Fig. 1e). These data indicated that the *hecw-1* 322C (in AB2 and CB3198) and *hecw-1* 325P (in CB4856) polymorphisms confer comparable effects on *P. aeruginosa* lawn avoidance behaviour, with delayed avoidance relative to the *hecw-1* 322Y 325Q allele crossed into the CB4856 background (Fig. 1e).

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

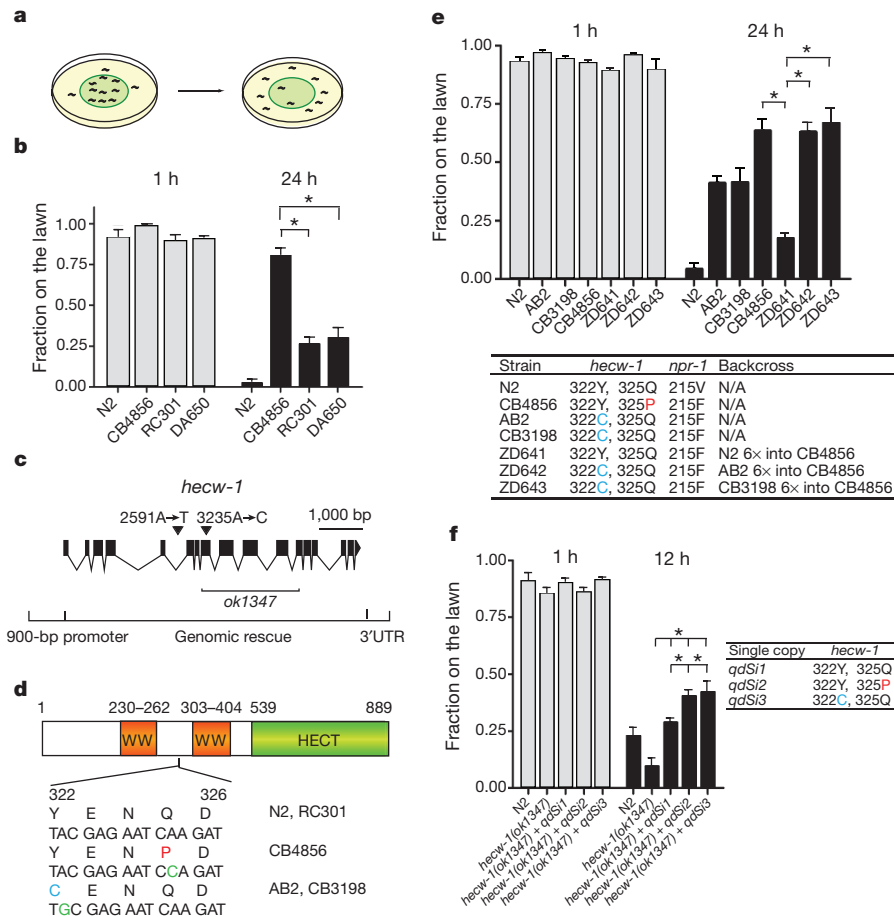


Figure 1 | Natural variation in *C. elegans* *hecw-1* modulates behavioural avoidance of *P. aeruginosa*. **a**, Schematic of the assay of pathogen avoidance behaviour in which *C. elegans* were transferred to plates containing a lawn of *P. aeruginosa* PA14, and lawn occupancy was scored over time. **b**, *P. aeruginosa* lawn occupancy of *C. elegans* strains assayed at 24 h. Full time course in Supplementary Fig. 1. **c**, Schematic of the 8-kb rescuing genomic fragment encompassing *hecw-1* (Supplementary Fig. 2c). The two polymorphisms that vary between CB4856 and DA650 within this region are shown, as is the region deleted in the *hecw-1(ok1347)* mutant. **d**, The coding polymorphisms in *hecw-1* result in a Q325P substitution in CB4856 and a Y322C substitution in the AB2

Although the strains carrying the 322C and 325Q alleles of *hecw-1* in the CB4856 background were generated from two independent strains, AB2 and CB3198, we considered that these strains may nevertheless carry linked loci that could modulate pathogen avoidance behaviour. Thus, we sought to determine the relative activities of the naturally occurring alleles of *hecw-1* in an otherwise isogenic strain background. We generated single-copy insertions each carrying an 8-kb *hecw-1* genomic sequence fragment in the N2-derived *hecw-1(ok1347)* background, with the transgenes differing only at HECW-1 amino acid positions 322(Y/C) and/or 325(Q/P). Consistent with our observations of strains carrying the naturally occurring *hecw-1* polymorphisms in the CB4856 background, we observed that animals carrying either the *hecw-1* 322C or the *hecw-1* 325P substitutions rescued the *hecw-1(ok1347)* *P. aeruginosa* lawn avoidance phenotype with a delayed avoidance of the *P. aeruginosa* lawn compared with animals carrying the *hecw-1* 322Y 325Q allele (Fig. 1f). These data show that both the *hecw-1* 322C 325Q and 322Y 325P polymorphisms result in increased HECW-1 activity relative to the *hecw-1* 322Y 325Q allele. Structural modelling of HECW-1, based on the crystal structure of the corresponding domain of human HECW1, indicated that the *hecw-1* 325P and *hecw-1* 322C polymorphisms may alter the same

and CB3198 wild strains. **e**, *P. aeruginosa* lawn avoidance behaviour of strains carrying *hecw-1* 322Y 325Q, 322Y 325P and 322C 325Q alleles in the CB4856 background and constructed as indicated. **f**, *P. aeruginosa* lawn avoidance behaviour of animals each carrying a single integrated copy of *hecw-1* genomic sequence with the naturally occurring 322Y 325Q, 322C 325Q and 322Y 325P alleles of *hecw-1* in the N2-derived *hecw-1(ok1347)* background. In **b** and **e** * $P < 0.001$ was determined by the analysis of variance (ANOVA) multiple comparisons test. In **f** * $P < 0.001$ and ** $P < 0.02$ was determined by the Student's *t*-test based on the results of four independent experiments. Error bars indicate s.e.m.

protein–protein interaction interface on the surface of HECW-1 without a radical disruption in overall structure (Supplementary Fig. 4b).

To determine the cells in which HECW-1 is expressed, we generated a transcriptional reporter consisting of green fluorescent protein (GFP) under the control of the *hecw-1* promoter, which contains the 0.9 kb sequence upstream of the first exon of *hecw-1* that was sufficient to rescue the *hecw-1(ok1347)* mutant. GFP fluorescence was observed in the nervous system throughout the body, but in the anterior ganglion was conspicuously limited to two neurons located posterior to the anterior bulb of the pharynx, each of which have projections extending anteriorly (Fig. 2a). We confirmed that these two neurons are the lateral outer labial neurons OLL and OLLR by co-localization experiments using the OLL-specific reporter *ser-2dp::gfp²¹* and the *hecw-1p::mCherry* transgene (Fig. 2b–d).

To determine whether expression of HECW-1 in the OLL neuron pair is sufficient to regulate pathogen avoidance behaviour, we performed rescue experiments using transgenes comprised of *hecw-1* complementary DNA fused to GFP under the control of neuron-specific promoters. Expression of a *hecw-1* cDNA::gfp transgene under the control of either its endogenous *hecw-1* promoter (Supplementary Fig. 5) or the OLL-specific *ser-2d* promoter²¹ (Fig. 2e) was sufficient to

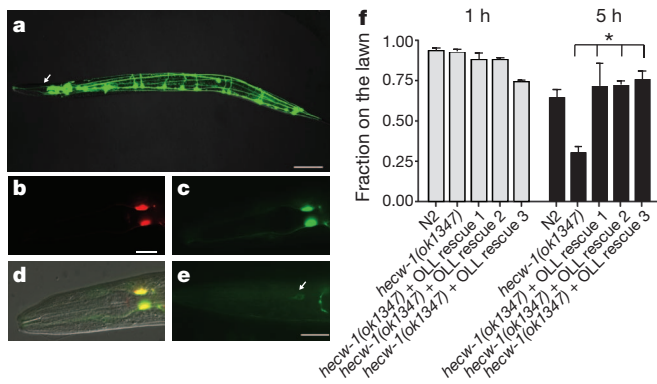


Figure 2 | HECW-1 functions in the OLL sensory neuron pair to negatively regulate pathogen avoidance behaviour. **a**, Expression of GFP under the control of a *hecw-1p* showing fluorescence throughout the nervous system of the body, but limited expression in the anterior ganglion (arrow points to single pair of head neurons showing expression). Scale bar, 50 μ m. **b–d**, HECW-1 is expressed in the OLL sensory neurons. **b**, Anterior ganglion expression of *hecw-1p::mCherry*. **c**, Expression of the OLL sensory neuron marker *ser-2dp::gfp*. **d**, Merge of Nomarski imaging, **b** and **c**. Scale bar, 5 μ m. **e**, Expression of a *ser-2dp::hecw-1 cDNA::gfp* translational fusion protein in the OLL neuron pair (indicated by arrow). Scale bar, 10 μ m. **f**, *P. aeruginosa* lawn avoidance behaviour of N2, *hecw-1(ok1347)* and three independent transgenic lines of the *hecw-1(ok1347)* mutant expressing *ser-2p::hecw-1::gfp*. * $P < 0.001$ was determined by the ANOVA multiple comparisons test. Error bars indicate s.e.m.

rescue the *hecw-1(ok1347)* *P. aeruginosa* lawn avoidance phenotype (Fig. 2f). Although the *ser-2d* promoter directs additional expression in PVD neurons²¹, heterologous expression of *hecw-1::gfp* in PVD did not rescue the *P. aeruginosa* lawn avoidance phenotype of *hecw-1* (Supplementary Fig. 6a). In addition, an alternative promoter directing *hecw-1::gfp* expression in neurons including the OLL neuron pair, but not PVD neurons, rescues the *hecw-1* mutant phenotype (Supplementary Fig. 6b). These data indicate that HECW-1 activity in the OLL neuron pair is sufficient to rescue the *P. aeruginosa* lawn avoidance phenotype of *hecw-1(ok1347)*.

To determine whether the OLL neuron pair is necessary for the negative regulation of pathogen avoidance behaviour, we carried out laser ablation of the OLL neuron pair using transgenic animals carrying the *ser-2dp::gfp* reporter to mark the OLL neurons for ablation. We ablated and mock-ablated the OLL neuron pair and allowed the animals to recover overnight on plates seeded with *E. coli* OP50. To confirm that the OLL neurons were successfully ablated, we checked the OLL GFP signals of the ablated animals before we transferred them to *P. aeruginosa* plates. We found that animals with the OLL neuron pair ablated showed markedly enhanced *P. aeruginosa* lawn avoidance behaviour relative to mock-ablated animals as determined by diminished occupancy of the *P. aeruginosa* lawn after 5 h (Fig. 3a). These data suggest a role for the OLL neuron pair in the negative regulation of pathogen avoidance behaviour in *C. elegans*.

We also generated transgenic animals in the N2 background expressing the caspase *csp-1b* cDNA (Denning, D. and Horvitz, H. R., unpublished results), which induces cell death, under the control of the OLL-specific promoter, *ser-2dp*. Using two transgenic lines with more than 90% of the animals lacking the OLL neuron pair (Fig. 3b, c), we found that animals lacking the OLL neuron pair showed an enhanced pathogen avoidance behaviour phenotype as was observed in OLL laser-ablated animals (Fig. 3d). The enhanced *P. aeruginosa* lawn avoidance behaviour of the *hecw-1(ok1347)* mutant and OLL-ablated strains led us to ask whether the lawn avoidance behaviour might be observed even in the presence of non-pathogenic *E. coli* OP50. However, in contrast to the marked avoidance behaviour observed in the presence of a lawn of pathogenic *P. aeruginosa*, we found that neither the *hecw-1(ok1347)* deletion, nor genetic ablation of OLL

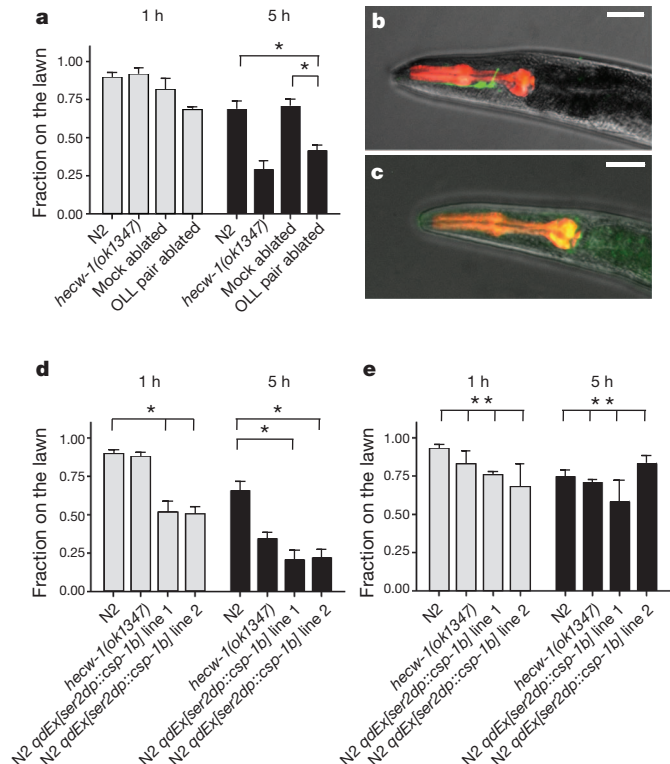


Figure 3 | The OLL sensory neuron pair negatively regulates pathogen avoidance behaviour. **a**, *P. aeruginosa* lawn avoidance behaviour of N2 animals carrying the *ser-2dp::gfp* transgene and subjected to laser ablation of OLL. Mock animals were treated in parallel without laser treatment. Two independent experiments were performed, and 15 and 30 animals were ablated in each experiment. **b**, **c**, Fluorescence microscopy of animals expressing the *ser-2dp::gfp* transgene in OLL (**b**) and carrying the *ser-2dp::gfp* transgene along with the *csp-1b* caspase under the control of *ser-2d* promoter that results in genetic ablation of OLL (**c**). Pharyngeal red fluorescence signals arise from co-injection marker *myo-2p::mcherry*. Scale bar, 20 μ m. **d**, **e**, Genetically ablated animals were then transferred to plates seeded with *P. aeruginosa* PA14 (**d**) and *E. coli* OP50 (**e**), respectively, and lawn occupancy was recorded at the indicated times. * $P < 0.001$ was determined by the ANOVA multiple comparisons test. ** $P = 0.23$ was determined by the ANOVA multiple comparisons test. Error bars indicate s.e.m.

mediated by *ser-2dp::csp-1b* conferred a discernible lawn avoidance behavioural phenotype in the presence of *E. coli* OP50 (Fig. 3e).

Ultrastructural studies suggest a mechanosensory function for the OLL neuron pair²². Consistent with these data, we found that OLL-ablated and *hecw-1(ok1347)* mutant animals showed a defective withdrawal response of *C. elegans* to a light touch on the tip of its nose, the nose-touch phenotype²³ (Supplementary Fig. 7a, b). The involvement of the OLL neuron pair in *P. aeruginosa* lawn avoidance behaviour suggests that the mechanosensory detection of bacteria may contribute to HECW-1-regulated pathogen avoidance behaviour. A role for mechanosensation in *C. elegans*–bacteria interactions has been implicated previously in studies of dopamine-dependent mechanosensory signalling in the *C. elegans* basal slowing response when encountering a lawn of bacteria²⁴.

The observed *P. aeruginosa* lawn avoidance phenotype probably results from the integration of multiple attractive and repulsive behaviours induced by nutritional, metabolic and pathogenic aspects of the bacterial lawn. We have previously shown that *P. aeruginosa* lawn avoidance behaviour of the *npr-1* mutant contributes to enhanced survival in a pathogen killing assay with *P. aeruginosa*^{11,14}. Consistent with these prior studies, we observed that the *hecw-1(ok1347)* mutant showed enhanced survival compared with N2 (Fig. 4a). We proceeded to ask whether HECW-1 might act through NPR-1 to modulate behavioural avoidance of pathogenic bacteria. Whereas the *hecw-1(ok1347)*

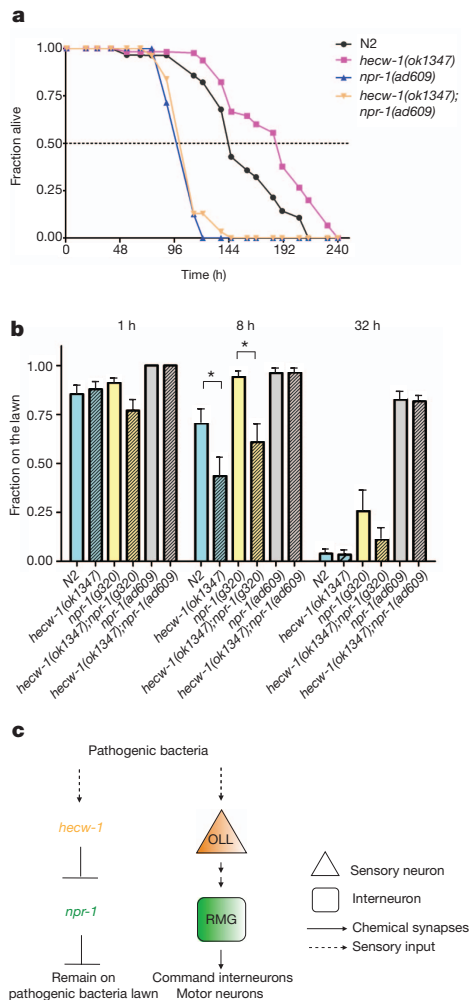


Figure 4 | Regulation of pathogen avoidance behaviour and survival by HECW-1 is dependent on NPR-1. **a**, Survival assay of indicated mutants on *P. aeruginosa* PA14 at 22.5 °C. **b**, *P. aeruginosa* lawn avoidance behaviour of double mutants in the N2 background carrying the *hecw-1(ok1347)* mutation in combination with three different *npr-1* alleles—the N2 wild type *npr-1* 215V allele, the 215F allele *npr-1(g320)*, and the putative null *npr-1(ad609)*. * $P < 0.001$ was determined by the ANOVA multiple comparisons test. Error bars indicate s.e.m. **c**, Model for the function of HECW-1 in the regulation of pathogen avoidance behaviour.

mutation conferred an enhanced *P. aeruginosa* lawn avoidance phenotype in the presence of the *npr-1* 215V allele and *npr-1(g320)* 215F alleles, we found that the putative null alleles, *npr-1(ad609)* and *npr-1(ky13)*, suppressed the enhanced *P. aeruginosa* avoidance phenotype conferred by the *hecw-1(ok1347)* mutation (Fig. 4b and Supplementary Fig. 8a). Consistent with the observed genetic interaction between *hecw-1* and *npr-1* with regard to *P. aeruginosa* lawn avoidance behaviour, we observed that both the pathogenesis survival and nose-touch phenotypes conferred by the *hecw-1* loss-of-function in the presence of *P. aeruginosa* were also suppressed by null mutations in *npr-1* (Fig. 4a and Supplementary Fig. 7b). The OLL neurons may connect through the IL2 and CEP neurons to the RMG neurons, in which NPR-1 activity exerts its diverse influence over *C. elegans* behaviour²⁵. Partial rescue of the sensitivity of the *npr-1(ky13)* mutant to the effect of the *hecw-1(ok1347)* mutation on pathogen avoidance was observed using an *npr-1* cDNA transgene under the control of the *flp-5* promoter, which directs expression in RMG and ASE neurons (Supplementary Fig. 8b). These data support a model in which HECW-1 in the OLL neuron pair functions to inhibit the activity of NPR-1 in the RMG inter/motor neuron (Fig. 4c).

The ecological and evolutionary impact that microbes may have on host organisms has been increasingly appreciated to extend beyond the immune system, encompassing diverse aspects of host physiology²⁶. Our work provides a molecular basis for how natural variation can lead to changes in behaviour that may facilitate adaptation of *C. elegans* to microbes.

METHODS SUMMARY

C. elegans were cultivated and strains constructed using standard methods²⁷. Transgenic strains were generated using indicated constructs and standard micro-injection methods²⁸. MosSCI single-copy insertions were generated as described²⁹. Assays measuring *C. elegans* pathogen avoidance behaviour used *P. aeruginosa* PA14 plates prepared as follows: a 100-ml solution of LB was inoculated with a single colony of *P. aeruginosa* PA14 and grown overnight without shaking at 37 °C ($A_{600\text{ nm}} = 0.2\text{--}0.3$). A 30- μl aliquot of this culture was used to seed the centre of each 100-mm NGM plate, and the seeded plates were incubated for 24 h at room temperature (22.5 °C). Approximately 30 larval stage 4 (L4) animals were transferred onto plates containing the *P. aeruginosa* PA14 lawn at 22.5 °C, and occupancy was determined at the indicated times. Survival assays were carried out on 35-mm slow-killing assay plates³⁰ supplemented with 5-fluorodeoxyuridine (0.05 mg ml⁻¹) and seeded with *P. aeruginosa* PA14 prepared as above and maintained at 22.5 °C. Laser ablations were performed on L3-stage larvae as described in text. Microscopy and image analysis was carried out on an AxioImager Z1 fluorescence microscope fitted with CCD camera (AxioCam) and processed with the Axioplan image processor software (Zeiss). The statistical analyses were performed using the GraphPad Prism software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 December 2010; accepted 17 October 2011.

Published online 16 November 2011.

- Flint, J. & Mackay, T. F. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
- Bendesky, A., Tsunozaki, M., Rockman, M. V., Kruglyak, L. & Bargmann, C. I. Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature* **472**, 313–318 (2011).
- Osborne, K. A. *et al.* Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**, 834–836 (1997).
- Barrière, A. & Félix, M. A. Natural variation and population genetics of *Caenorhabditis elegans* in WormBook (ed. The *C. elegans* Research Community) doi/10.1895/wormbook.1.43.1 (2005).
- Félix, M. A. & Braendle, C. The natural history of *Caenorhabditis elegans*. *Curr. Biol.* **20**, R965–R969 (2010).
- Kim, D. H. *et al.* A conserved p38 MAP kinase pathway in *Caenorhabditis elegans* innate immunity. *Science* **297**, 623–626 (2002).
- Mallo, G. V. *et al.* Inducible antibacterial defense system in *C. elegans*. *Curr. Biol.* **12**, 1209–1214 (2002).
- Nicholas, H. R. & Hodgkin, J. The ERK MAP kinase cascade mediates tail swelling and a protective response to rectal infection in *C. elegans*. *Curr. Biol.* **14**, 1256–1261 (2004).
- Zhang, Y., Lu, H. & Bargmann, C. I. Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*. *Nature* **438**, 179–184 (2005).
- de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).
- Reddy, K. C., Andersen, E. C., Kruglyak, L. & Kim, D. H. A polymorphism in *npr-1* is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science* **323**, 382–384 (2009).
- Pradel, E. *et al.* Detection and avoidance of a natural product from the pathogenic bacterium *Serratia marcescens* by *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **104**, 2295–2300 (2007).
- Pujol, N. *et al.* A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. *Curr. Biol.* **11**, 809–821 (2001).
- Shivers, R. P., Kooistra, T., Chu, S. W., Pagano, D. J. & Kim, D. H. Tissue-specific activities of an immune signaling module regulate physiological responses to pathogenic and nutritional bacteria in *C. elegans*. *Cell Host Microbe* **6**, 321–330 (2009).
- McGrath, P. T. *et al.* Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* **61**, 692–699 (2009).
- Weber, K. P. *et al.* Whole genome sequencing highlights genetic changes associated with laboratory domestication of *C. elegans*. *PLoS ONE* **5**, e13922 (2010).
- Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* **4**, e274 (2006).
- Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).

19. Reddy, K. C., Hunter, R. C., Bhatla, N., Newman, D. K. & Kim, D. H. *Caenorhabditis elegans* NPR-1-mediated behaviors are suppressed in the presence of mucoid bacteria. *Proc. Natl Acad. Sci. USA* **108**, 12887–12892 (2011).
20. Miyazaki, K. *et al.* NEDL1, a novel ubiquitin-protein isopeptide ligase for dishevelled-1, targets mutant superoxide dismutase-1. *J. Biol. Chem.* **279**, 11327–11335 (2004).
21. Tsalik, E. L. *et al.* LIM homeobox gene-dependent expression of biogenic amine receptors in restricted regions of the *C. elegans* nervous system. *Dev. Biol.* **263**, 81–102 (2003).
22. Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
23. Kaplan, J. M. & Horvitz, H. R. A dual mechanosensory and chemosensory neuron in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 2227–2231 (1993).
24. Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
25. Macosko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009).
26. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
27. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
28. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
29. Frøkjær-Jensen, C. *et al.* Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genet.* **40**, 1375–1383 (2008).
30. Tan, M. W., Mahajan-Miklos, S. & Ausubel, F. M. Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proc. Natl Acad. Sci. USA* **96**, 715–720 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Denning and H. R. Horvitz for the *csp-1b* cDNA. We thank Y. Kohara for the *hecw-1* and *npr-1* cDNAs. We thank J. Meisel for sequencing of *hecw-1* polymorphisms. We thank D. Ma for technical advice on laser ablation. We thank T. Schwartz for expert advice with the structural modeling of HECW-1. The *hecw-1(ok1347)* deletion allele was generated by the *C. elegans* Knockout Consortium and obtained, along with other strains used in this study, from the *Caenorhabditis* Genetics Center (CGC), which is supported by the NIH—National Center for Research Resources. This work was supported by NIH Grant GM084477 (to D.H.K.).

Author Contributions H.C.C. and D.H.K. designed experiments, H.C.C. and J.P. performed experiments, H.C.C. and D.H.K. analyzed and interpreted the data, and H.C.C. and D.H.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.H.K. (dhkim@mit.edu).

METHODS

Strains. *C. elegans* strains N2 (Bristol, England), CB4856 (Hawaii, USA), RC301 (Freiburg, Germany), AB2 (Adelaide, Australia) and CB3198 (Pasadena, California, USA) were maintained at 20 °C using standard methods²⁷. Strains were maintained at 20 °C then shifted to 22.5 °C for *P. aeruginosa* PA14 lawn avoidance assays and pathogenesis survival assays. The *hecw-1(ok1347)* mutant strain was derived from the RB1265 strain (generated by the *C. elegans* Gene Knockout Consortium and obtained from the *Caenorhabditis* Genetics Center) by backcrossing six times to N2. The *npr-1(g320); hecw-1(ok1347)* double mutant was generated by crossing *hecw-1(ok1347)* into DA650 three times. ZD641, ZD642, ZD643 were generated by backcrossing the *hecw-1* alleles in N2, AB2 and CB3198 into CB4856 six times. MT15623 (a gift from H. R. Horvitz) carries *npr-1(ky13); lin-15AB(n765ts)* X; *nEx1252 (npr-1* genomic; *lin-15AB* rescue) were used and crossed into *hecw-1(ok1347); npr-1(ky13); lin-15AB(n765ts)* for rescue experiments. Transgenic strains were isolated by microinjecting various plasmids²⁸, (typically at 50–100 µg ml⁻¹) together with either of the following co-injecting markers *rol-6dm*, *myo-2p::mCherry* (a gift from S. Nakano) and *mec-4p::gfp* (a gift from M. Driscoll) in wild-type or mutant animals.

Generation of MosSCI single insertion lines. Single-copy insertion of transgenes was performed using the direct MosSCI technique targeting the *tTt5605* Mos allele on chromosome II, as described²⁹. The *hecw-1* genomic sequences varying only in the coding polymorphisms in exon 8 were cloned into the pCFJ151 targeting vector using the BsiWI and XhoI restriction sites. The 8-kb N2 *hecw-1* genomic sequences were modified by site-directed mutagenesis (QuikChange II XL Site-Directed Mutagenesis kit, Agilent Technologies) using the primer 5'-CTTGATTCTCGCATAAGGCGGTTTAACTCGTCTCG-3' to introduce the Q325P change and the 5'-CTGCATTGCATCTGGATTCTCGTATAAGGCGG-3' primer to introduce the Y322C change. The modified pCFJ151 vectors, carrying the *hecw-1* 322Y 325Q, *hecw-1* 322Y 325P or *hecw-1* 322C 325Q alleles, were then each injected with co-injection vectors pGH8, pCFJ90, pCFJ104 and pJL43.1 into at least 60 animals. After positive and negative selections, two integrated lines were then sequenced to confirming the correct sequences of *hecw-1*.

***P. aeruginosa* PA14 avoidance assay.** A 100-ml solution of LB was inoculated with a single colony of *P. aeruginosa* PA14 and grown overnight without shaking at 37 °C ($A_{600} = 0.2$ – 0.3). 30 µl of this culture was used to seed the centre of each 100-mm NGM plate, and the seeded plates were incubated for 24 h at room temperature (22.5 °C). Approximately 30 larval stage 4 (L4) animals were transferred onto plates containing the *P. aeruginosa* PA14 lawn at 22.5 °C, and lawn occupancy was measured at the indicated times. Two plates of each genotype were performed in each experiment and all experiments were performed at least three times. Upon being transferred to the *P. aeruginosa* PA14-containing plates, animals explore the plate for about 10–15 min until they find the bacterial lawn and then remain in the lawn. Subsequently, lawn occupancy is measured over time as the lawn avoidance behaviour is observed.

***P. aeruginosa* pathogenesis survival assay.** *C. elegans* survival assays on *P. aeruginosa* PA14 were set up as described previously³⁰, with the following modifications. 3 µl of the aforementioned *P. aeruginosa* PA14 culture was used to seed 35-mm slow-killing assay plates³⁰ modified by the addition of 5-fluorodeoxyuridine (0.05 mg ml⁻¹), used to prevent eggs from hatching. Seeded plates were incubated for 24 h at 22.5 °C. L4 stage animals were then transferred to the *P. aeruginosa* PA14 plates, which were maintained at 22.5 °C throughout the survival assay. Animals were scored as dead when they no longer responded to the repeated touch of a platinum wire.

Mapping of *hecw-1*. Single nucleotide polymorphisms (SNPs) between N2 and CB4856 were used to determine the genetic loci responsible for the difference in pathogen avoidance behaviour between DA650 and CB4856. We found a strong linkage in the area between -7.2 and -5.5 map units (centimorgan, cM) on chromosome 3. Further positional SNP walking defined a 25-kb interval between SNPs haw42334 (-7.18 cM) and haw42350 (-7.07 cM) on chromosome 3, which included the major molecular determinant of the *P. aeruginosa* PA14 lawn avoidance behaviour between CB4856 and DA650 (Supplementary Fig. 2). An 8-kb genomic region from CB4856 encompassing *hecw-1* and depicted in Fig. 1c rescued the DA650 avoidance phenotype.

Microscopy. Animals were mounted in M9 with levamisole (10 mM) onto slides with a 2% agarose pad. The slides were viewed using an AxioImager Z1 fluorescence microscope (Zeiss) with $\times 20/0.8$, $\times 40/1.3$ (oil) and $\times 63/1.4$ (oil) objectives. The fluorescence signals were recorded by a CCD camera (AxioCam) in a 12-bit format without saturation. The images were captured and processed using the Axioplan image processor software.

Molecular cloning. The genomic region of *hecw-1* was amplified by PCR using primers 5'-ATCGTGGTTTCGCACCTTTGTTTGGAAAC-3' and 5'-AACCAGTCGCCGTGAGGATGCG-3'. The 8 kb genomic fragment from either N2 or CB4856 was subsequently cloned into a Topo (Invitrogen) vector. The *hecw-1* promoter region was amplified by using 5' primer 5'-ATCGTGGTTTCGCACCTTTGTTTGGAAAC-3' and 3' primer 5'-TCTCCGAAGCGTGATCATCTCTAAAAACC-3'. The 900-bp fragment was subsequently cloned into the pPD95.75 vector (Fire Lab Vector Kit, Addgene) to generate the *hecw-1* promoter GFP reporter. *hecw-1* and *npr-1* cDNAs were gifts from Y. Kohara. The *ser-2d* promoter was generated using primers as described previously. *csp-1b* cDNA is a gift from D. Denning (D. Denning and H. R. Horvitz, unpublished results). *flp-5*, *flp-3* and F49H12.4 promoters were cloned from genomic DNA as described previously. The *flp-5* promoter region was amplified by using 5' primer 5'-TTTTTGCATGC GTCATGTCAGGCCCGCCGTTTC-3' and 3' primer 5'-TTTTTCTGCACTATTATGTGGAGCGGTTGGGAGCTG-3'. The *flp-3* promoter region was amplified by using 5' primer 5'-TTTTTCTGCAGACAAGGTGGCATCTCAACCAACAG-3' and 3' primer 5'-AAAAACCCGGGGAA GAAGTAGAAGTGAGGTTGTGAACCG-3'. Finally, the F49H12.4 promoter region was amplified by using 5' primer 5'-TTTTTCTGCAGCCTCTGTATGCCAGTTGGGGTTCTTGCT-3' and 3' primer 5'-AAAAACCCGGG GCAGCTGGAGGCTGGATATCTGAGG-3'.

Laser ablation. Laser microsurgery was performed on L3-stage larvae as described in text. Cell identifications were based on a *ser-2dp::gfp* and *sra-6p::gfp* reporter and cell morphology. The ablated animals were confirmed under a dissection scope to have lost the GFP signals of OLL and ASH neurons 16 h later, before the *P. aeruginosa* PA14 lawn avoidance and nose-touch assays were performed.

Nose-touch behaviour assay. Nose-touch sensory responses were assayed as previously described²³ in an observer-blinded manner. Each animal was tested on food for reversal of locomotion after a forward collision with a hair. Each animal was tested ten times, and 20 or more animals were tested for each genotype or for neuronal laser ablations.

Statistical analysis. Statistical analysis was performed using GraphPad Prism software.

Structure modelling. Based on the deposited WW domain of the human orthologue HECW1 (PDB code 3L4H), we threaded the *C. elegans* sequence (29% identity over 109 aligned residues) and generated a model with MODELLER. The model in Supplementary Fig. 4b is rendered using PyMOL.

Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*

Cécile Crosnier^{1*}, Leyla Y. Bustamante^{2*}, S. Josefin Bartholdson^{1*}, Amy K. Bei³, Michel Theron², Makoto Uchikawa⁴, Souleymane Mboup⁵, Omar Ndir⁵, Dominic P. Kwiatkowski^{2,6}, Manoj T. Duraisingh³, Julian C. Rayner² & Gavin J. Wright^{1,2}

Erythrocyte invasion by *Plasmodium falciparum* is central to the pathogenesis of malaria. Invasion requires a series of extracellular recognition events between erythrocyte receptors and ligands on the merozoite, the invasive form of the parasite. None of the few known receptor–ligand interactions involved^{1–4} are required in all parasite strains, indicating that the parasite is able to access multiple redundant invasion pathways⁵. Here, we show that we have identified a receptor–ligand pair that is essential for erythrocyte invasion in all tested *P. falciparum* strains. By systematically screening a library of erythrocyte proteins, we have found that the Ok blood group antigen, basigin, is a receptor for PfRh5, a parasite ligand that is essential for blood stage growth⁶. Erythrocyte invasion was potently inhibited by soluble basigin or by basigin knockdown, and invasion could be completely blocked using low concentrations of anti-basigin antibodies; importantly, these effects were observed across all laboratory-adapted and field strains tested. Furthermore, Ok^{a-} erythrocytes, which express a basigin variant that has a weaker binding affinity for PfRh5, had reduced invasion efficiencies. Our discovery of a cross-strain dependency on a single extracellular receptor–ligand pair for erythrocyte invasion by *P. falciparum* provides a focus for new anti-malarial therapies.

Among the many *P. falciparum* merozoite proteins that are believed to have a role in erythrocyte invasion, most attention has focused on two major parasite protein families: the EBAs and Rh^s. Although erythrocyte receptors have been identified for some of them (members of the glycophorin family are receptors for three EBAs^{1–3} and complement receptor 1 (CD35) has recently been identified as a receptor for PfRh4, ref. 4) none of these receptor–ligand pairs are essential in all parasite strains tested. PfRh5 is unique amongst the EBAs and Rh^s because it cannot be deleted in any *P. falciparum* strain and is therefore apparently essential for parasite growth in blood stage culture^{5,6}. Both native and recombinant PfRh5 have been previously shown to bind erythrocytes through an unknown glycosylated receptor that is resistant to chymotrypsin, trypsin and neuraminidase treatment^{6,8,9}.

To identify an erythrocyte receptor for PfRh5, we used a systematic screening approach by first compiling a library of abundant cell surface and secreted proteins expressed by human erythrocytes based on published proteomics data¹⁰. Proteins for which the entire ectodomain was expected to be expressed as a soluble recombinant protein were selected (Supplementary Table 1), and expressed by mammalian cells (Supplementary Fig. 1). The 40 proteins within the erythrocyte ectodomain protein library were then systematically screened using the AVExis assay (avidity-based extracellular interaction screen)¹¹ for interactions with a recombinant PfRh5 protein, also produced by mammalian cells. The AVExis assay is designed to detect direct low-affinity protein interactions between ectodomain fragments expressed as either biotin-tagged baits or highly avid pentameric β -lactamase-tagged

preys^{12,13}. The PfRh5 prey interacted with a single erythrocyte receptor bait (Fig. 1a, top panel) corresponding to the Ok blood group antigen, basigin (BSG, also known as CD147, EMMPRIN and M6, ref. 14). The same single interaction was identified in the reciprocal bait–prey orientation (Fig. 1a, lower panel).

BSG is a member of the immunoglobulin superfamily (IgSF) and has been implicated in many biological functions including embryo implantation, spermatogenesis¹⁵ and retinal development¹⁶. BSG exists in both long (three IgSF domains, BSG-L) and short (two IgSF domains, BSG-S) splice isoforms (Fig. 1b) and although BSG-L was used in the screen, BSG-S is thought to be the major isoform expressed on erythrocytes. Binding experiments using domain deletions established that PfRh5 could interact with BSG-S and this required both domains because neither of the two BSG-S IgSF domains were individually able to bind PfRh5 (Fig. 1b and Supplementary Fig. 2). We showed that PfRh5 interacted directly with BSG-S and BSG-L using purified proteins and surface plasmon resonance (SPR). Both kinetic (Fig. 1c) and equilibrium (Supplementary Fig. 3) binding parameters for the interaction were derived using a 1:1 binding model and were in excellent agreement (Supplementary Table 2). These parameters are typical of extracellular protein interactions measured using this technique¹⁷. Removal of glycans from BSG either by mutating all predicted glycosylation motifs or by enzymatic treatment did not affect PfRh5 binding (Supplementary Fig. 4), indicating that the PfRh5 binding site is solely located in the BSG protein core. BSG is also known to be resistant to trypsin and chymotrypsin treatment¹⁸, consistent with previous PfRh5–erythrocyte binding studies^{6,8,9}.

To determine whether the PfRh5–BSG interaction was required for invasion, we added purified pentamerized soluble BSG-S into invasion assays to specifically compete with the membrane-bound receptor. We found that BSG-S strongly inhibited invasion in a dose-dependent manner relative to controls which included each of the two non-binding BSG-S IgSF domains added individually (Fig. 2a). Strong inhibition was also observed across multiple strains (Fig. 2b) or when soluble BSG-L was added (Supplementary Fig. 5), although this was slightly weaker for the 3D7 strain. Soluble forms of BSG consisting of the extracellular regions are known to have biological effects such as upregulation of matrix metalloproteases¹⁹. To rule out an indirect effect of exogenous BSG on invasion, we added to invasion assays two independent purified anti-BSG monoclonal antibodies (MEM-M6/6 and TRA-1-85) which could both block the PfRh5–BSG interaction *in vitro* (data not shown). These high-affinity reagents gave a potent invasion blocking effect that was saturable at very low antibody concentrations (half-maximum inhibitory concentration $\approx 0.5 \mu\text{g ml}^{-1}$), consistent with binding and occluding a specific surface receptor of typical abundance ($\sim 10^4$ to 10^6 molecules per cell²⁰) (Fig. 2c). Pre-adsorption of the MEM-M6/6 antibody with soluble monomeric BSG specifically relieved the inhibition, ruling out any indirect effect of the antibody on non-BSG targets;

¹Cell Surface Signalling Laboratory, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ²Malaria Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.

³Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁴Tokyo Red Cross Blood Center, Tokyo 135-8639, Japan. ⁵Laboratory of Bacteriology and Virology, Le Dantec Hospital and Laboratory of Parasitology, Cheikh Anta Diop University, BP: 7325, Dakar, Senegal. ⁶Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.

*These authors contributed equally to this work.

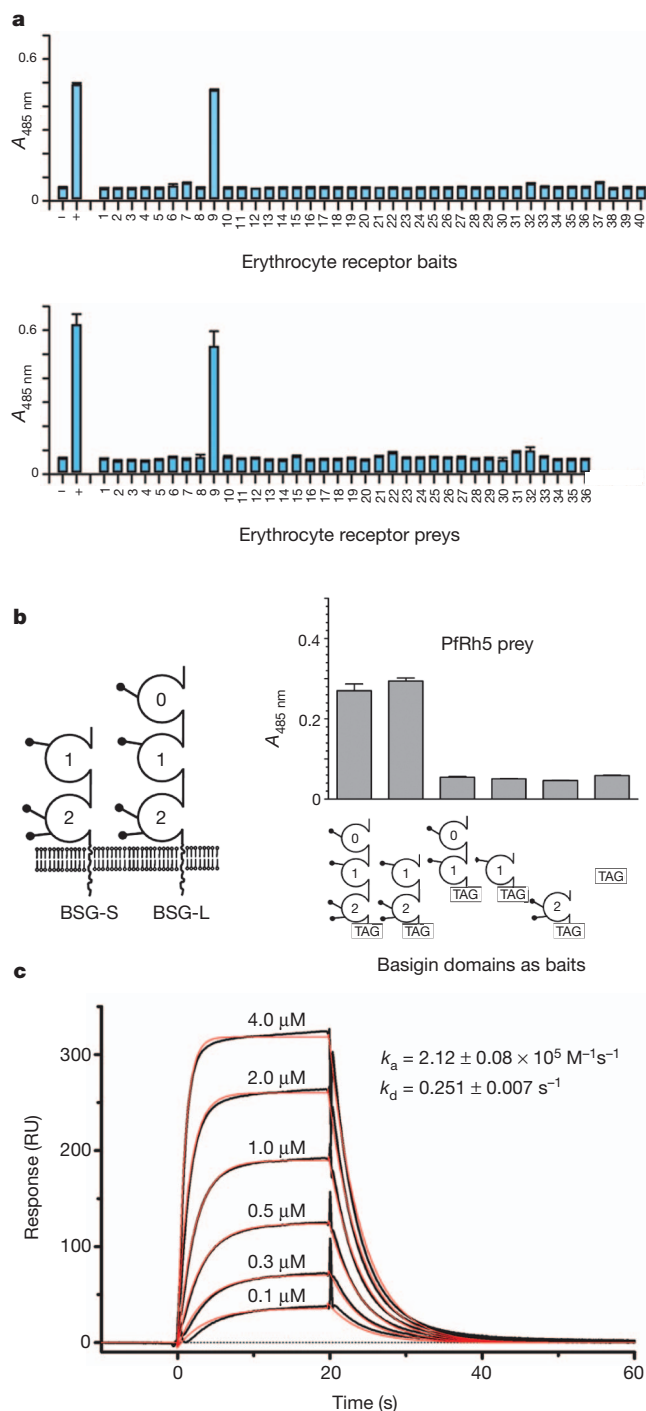


Figure 1 | BSG is an erythrocyte receptor for PfRh5. **a**, PfRh5 was screened as either a prey (top panel) or a bait (bottom panel) against an erythrocyte receptor protein library using AVEIXIS. BSG (protein 9) was identified as a receptor for PfRh5 in both bait–prey orientations. **b**, Domain structure of the BSG isoforms (left); lollipops represent potential N-linked glycosylation sites. BSG regions were expressed as baits and used to map the PfRh5 binding site to the two membrane-proximal domains. Bar charts show mean \pm s.e.m.; $n = 3$. **c**, Biophysical analysis of the PfRh5–BSG-S interaction using SPR. The indicated concentrations of purified PfRh5 were injected over immobilised BSG, and biophysical parameters derived from a 1:1 binding model (red line). RU, response units.

furthermore, MEM-M6/6 did not affect intra-erythrocytic *P. falciparum* development (Supplementary Fig. 6). Invasion was quantified using flow cytometry and a fluorescent DNA dye to stain parasites²¹. Using this assay, apparent invasion could not be eliminated, with efficiencies

reduced to a maximum of 80–90%, even at much higher concentrations of antibody (up to 1.5 mg ml^{-1} of MEM-M6/6, data not shown); however, direct observation of parasites using Giemsa-stained thin smears revealed that this residual staining in cytometry assays was due to extracellular parasites and debris in the culture. Using microscopy-based assays, we found that MEM-M6/6 concentrations of $10 \mu\text{g ml}^{-1}$ or more was sufficient to prevent all detectable invasion (Fig. 2d).

P. falciparum isolates can vary widely in their ability to invade erythrocytes treated with different receptor-modifying enzymes such as trypsin, chymotrypsin and neuraminidase, showing differential dependencies on erythrocyte receptors for invasion. To determine if BSG was a critical invasion receptor across *P. falciparum* lines that use different invasion pathways, we tested the ability of MEM-M6/6 to block erythrocyte invasion on nine culture-adapted strains representing seven different PfRh5 sequence variants (Supplementary Table 3). We observed that the invasion of all lines was potently inhibited by MEM-M6/6 (Fig. 2e). To show that the dependency on BSG was not an unusual feature of culture-adapted lines, we also tested six freshly-isolated *P. falciparum* strains from Senegal²² and again observed a potent inhibitory effect (Fig. 2f). Assays with the field isolates were carried out with unsynchronised parasites, decreasing the overall inhibitory effect because not all parasites had reinvasion over the course of the assay. All six Senegal isolates, however, were inhibited by MEM-M6/6 to the same extent as an unsynchronised culture-adapted line, W2mef, tested at the same time. This demonstrated that freshly-isolated field strains have the same dependency on BSG as laboratory-adapted lines (Fig. 2f).

To confirm independently the essentiality of BSG as a *P. falciparum* invasion receptor, we used a genetic approach by differentiating erythrocytes from haematopoietic stem cells transduced with lentiviruses containing either a short hairpin RNA targeting BSG or a scrambled shRNA control (pLKO). BSG-targeted erythrocytes showed a reproducible knockdown to approximately 50 to 60% of cell surface BSG levels relative to the pLKO control (Fig. 2g) and expressed markers indicative of complete erythrocyte maturation (Supplementary Fig. 7). The invasion of both the 3D7 and W2mef *P. falciparum* strains into BSG-knockdown erythrocytes was significantly reduced compared to the control (18% versus 94% for 3D7 and 14% versus 103% for W2mef, Fig. 2h). By contrast, previous knockdown of GYPA, the major surface sialoglycoprotein, significantly inhibited the W2mef but not the 3D7 strain²³. The inhibition of erythrocyte invasion by multiple *P. falciparum* strains using soluble BSG, anti-BSG monoclonal antibodies, or knockdown of BSG surface expression suggests that BSG is a critical host receptor for *P. falciparum* invasion.

Malaria is thought to have been a strong selective pressure in human evolutionary history and given the apparently essential roles of PfRh5 and BSG in *P. falciparum* invasion we sought to determine if any human populations contained genetic variants in BSG that might affect PfRh5 binding and invasion. Five nonsynonymous single nucleotide polymorphisms (SNPs) have been described within the BSG-S IgSF domains (Supplementary Table 4 and Fig. 3a). These variants were expressed and the biophysical PfRh5 binding parameters determined using SPR. Equilibrium measurements showed that two variants had lower binding affinity compared to the BSG reference sequence: L90P and E92K (Fig. 3b and Supplementary Table 2). L90P did not interact with PfRh5 and binding profiles of several anti-BSG monoclonal antibodies indicated local misfolding of the membrane-distal IgSF domain (Supplementary Fig. 8). No verification or population frequency data for this SNP are currently available, preventing further biological interpretation of this variant. E92K had a twofold lower affinity for PfRh5 (Fig. 3b) and a comparative kinetic analysis demonstrated that this was due to both a slower association and a faster dissociation rate (Fig. 3c and Supplementary Table 2). The E92 residue is solvent-exposed and located within the loop connecting

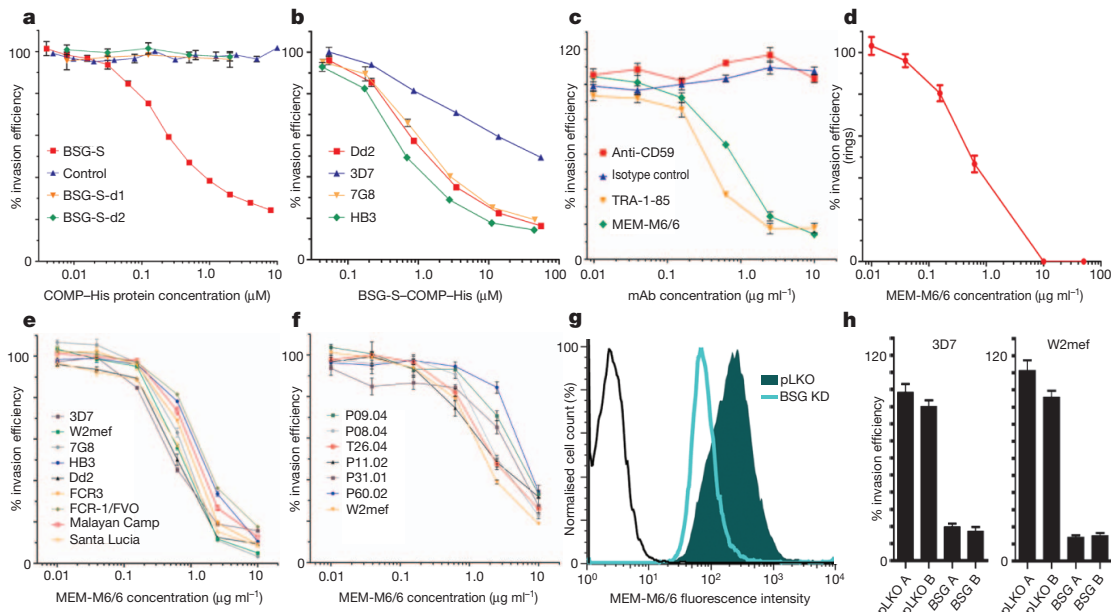


Figure 2 | Soluble BSG, anti-BSG antibodies and BSG knockdown potently block erythrocyte invasion. **a**, Erythrocyte invasion was inhibited by purified pentamerized BSG-S-Cd4d3+4-COMP-His ectodomains but not by the two non-binding BSG-S domains added individually or Cd4d3+4-COMP-His (control); strain Dd2. **b**, Cross-strain inhibition of invasion using pentamerized BSG-S. **c**, Anti-BSG monoclonal antibodies, TRA-1-85 and MEM-M6/6, potently inhibited invasion of erythrocytes; strain 3D7. **d**, MEM-M6/6 concentrations $\geq 10 \mu\text{g ml}^{-1}$ prevented all detectable invasion by microscopic

observation of cultures; strain 3D7. **e**, **f**, MEM-M6/6 inhibited invasion of synchronised *P. falciparum* culture-adapted lines (**e**) and unsynchronised field isolates (**f**). **g**, Cell-surface BSG is reduced in erythrocytes differentiated from haematopoietic stem cells transduced with lentiviruses containing shRNA targeting BSG (light blue line) relative to a control virus (pLKO, shaded); black line represents secondary antibody alone. **h**, 3D7 and W2mef invasion was inhibited in BSG knockdown erythrocytes. A and B are replicates. Invasion efficiencies are mean \pm s.e.m., $n = 3$.

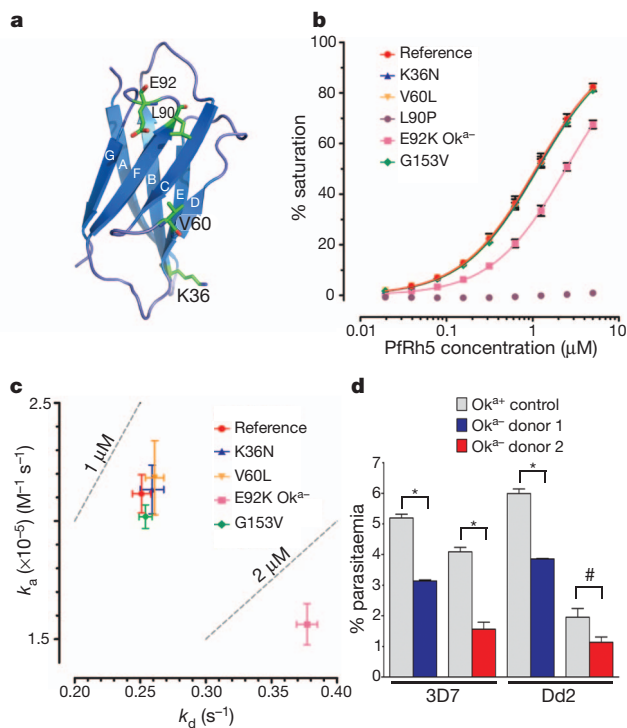


Figure 3 | The Ok^{a-} BSG variant has reduced binding affinity for PfRh5 and Ok^{a-} erythrocytes have reduced merozoite invasion frequencies. **a**, Schematic of the membrane distal IgSF domain of BSG-S showing the location of naturally occurring variants. **b**, Equilibrium binding isotherms of PfRh5 binding to BSG-S variants. **c**, Association (k_a) and dissociation (k_d) rate constants of PfRh5 binding to BSG-S and variants. Means \pm s.e.m.; $n = 3$. **d**, Invasion of 3D7 and Dd2 strains in Ok^{a-} blood cells are reduced relative to the Ok^{a+} control. Mean \pm s.e.m., $n = 3$; * $P \leq 0.0003$; # $P = 0.0349$, unpaired one-tailed *t* test. A repeat is shown in Supplementary Fig. 9.

the F-G β -strands close to the glycan-free GFC β -sheet, consistent with a possible PfRh5-binding interface (Fig. 3a). E92K is the variant responsible for the Ok^{a-} blood group, which has been described in eight Japanese families¹⁴. Ok^{a-} erythrocytes from two unrelated donors showed reduced invasion with both 3D7 and Dd2 *P. falciparum* strains relative to Ok^{a+} controls (Fig. 3d and Supplementary Fig. 9), correlating with the reduced affinity of the Ok^{a-} variant for PfRh5. The extreme rarity and restriction of the Ok^{a-} blood group to Japanese individuals suggest that this specific allele has not had a major role in conferring resistance to malaria. It is possible that other BSG polymorphisms, as yet unknown, have evolved in some malaria-exposed populations as a mechanism of resistance to *P. falciparum*. The search for functional polymorphisms of BSG needs to go beyond gene coding regions as the results of our knockdown experiments indicate that expression levels of BSG at the erythrocyte surface influence the ability of the parasite to invade. The Duffy variant, which confers resistance to *P. vivax*, is also a non-coding regulatory polymorphism that suppresses expression of the invasion receptor by erythrocytes. Our ability to address this problem is currently limited by the lack of data on genome variation among the many different ethnic groups that are exposed to *P. falciparum* malaria, but will be greatly enhanced by the 1000 Genomes Project, MalariaGEN and other genetic studies that are now in progress in Africa and other malaria-endemic regions of the world^{24–26}. Inter-population comparisons of haplotype length and frequency provide a potentially powerful way of addressing this problem²⁷, and there is preliminary evidence that a region of chromosome 19 encompassing BSG and several neighbouring genes has undergone recent positive selection in West Africa, but a considerable amount of further work is needed to determine whether this is causally related to the role of BSG as a malaria invasion receptor (MalariaGEN consortium, unpublished data).

In summary, we have applied a systematic protein interaction screening approach (AVEXIS) to identify BSG as an erythrocyte receptor for PfRh5. Importantly, we were able to prevent all detectable erythrocyte invasion by every *P. falciparum* strain that we tested using only modest concentrations of anti-BSG antibodies. These

observations, coupled with the inability to delete *PfRh5*⁶, lead us to conclude that the interaction between BSG and PfRh5 is essential for parasite entry, and may perform a fundamentally different function to the other EBA and Rh proteins, which are involved in redundant, partially overlapping invasion pathways. The dependence on a single receptor–ligand pair across many *P. falciparum* strains may provide new possibilities for therapeutic intervention.

METHODS SUMMARY

Recombinant protein production and interaction screening. Protein production, purification, AVExis assays and SPR were performed essentially as described¹¹ except the type II proteins which were expressed with an amino-terminal C4d3+4-biotin tag and a mouse antibody signal peptide. PfRh5 was expressed as above except that a non-endogenous signal peptide was added and the threonines in potential N-linked glycan sequons were mutated to alanine to prevent inappropriate glycosylation. All constructs were chemically synthesized and codon optimised for mammalian expression (GeneArt AG). Purified pentameric proteins used in invasion assays were made by replacing the β -lactamase reporter in the prey plasmid with a hexa-his tag, purified and buffer exchanged into RPMI before use. BSG variants were produced by site directed mutagenesis.

***P. falciparum* culture, lentiviral transduction and invasion assays.** *P. falciparum* parasite strains were routinely cultured in human O+ erythrocytes at 5% haematocrit. Use of erythrocytes from human donors for *P. falciparum* culture was approved by the NHS Cambridgeshire 4 Research Ethics Committee. Ok^{a+} and control Ok^{a+} blood was obtained from donors in Japan with informed consent, shipped on ice and experiments performed within 72 h. Invasion assays were carried out as described previously²¹ using the two-colour assay for the Ok^{a+} experiment. Lentiviral transduction of HSCs was performed as previously described²³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 March; accepted 3 October 2011.

Published online 9 November 2011; corrected 21 December 2011 (see full-text HTML version for details).

- Maier, A. G. *et al.* *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nature Med.* **9**, 87–92 (2003).
- Mayer, D. C. *et al.* Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proc. Natl Acad. Sci. USA* **106**, 5348–5352 (2009).
- Sim, B. K. *et al.* Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* **264**, 1941–1944 (1994).
- Tham, W. H. *et al.* Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proc. Natl Acad. Sci. USA* **107**, 17327–17332 (2010).
- Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–766 (2006).
- Baum, J. *et al.* Reticulocyte-binding protein homologue 5 – an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *Int. J. Parasitol.* **39**, 371–380 (2009).
- Iyer, J. *et al.* Invasion of host cells by malaria parasites: a tale of two protein families. *Mol. Microbiol.* **65**, 231–249 (2007).
- Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4**, 40–51 (2008).
- Rodriguez, M. *et al.* PfRH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE* **3**, e3300 (2008).
- Pasini, E. M. *et al.* In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **108**, 791–801 (2006).
- Bushell, K. M. *et al.* Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* **18**, 622–630 (2008).
- Martin, S. *et al.* Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol. Cell. Proteomics* **9**, 2654–2665 (2010).
- Söllner, C. & Wright, G. J. A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biol.* **10**, R99 (2009).
- Spring, F. A. *et al.* The Ok^a blood group antigen is a marker for the M6 leukocyte activation antigen, the human homolog of OX-47 antigen, basigin and neurothelin, an immunoglobulin superfamily molecule that is widely expressed in human cells and tissues. *Eur. J. Immunol.* **27**, 891–897 (1997).
- Igakura, T. *et al.* A null mutation in basigin, an immunoglobulin superfamily member, indicates its important roles in peri-implantation development and spermatogenesis. *Dev. Biol.* **194**, 152–165 (1998).
- Fadool, J. M. & Linser, P. J. 5A11 antigen is a cell recognition molecule which is involved in neuronal-glial interactions in avian neural retina. *Dev. Dyn.* **196**, 252–262 (1993).
- Wright, G. J. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. Biosyst.* **5**, 1405–1412 (2009).
- Williams, B. P. *et al.* Biochemical and genetic analysis of the Ok^a blood group antigen. *Immunogenetics* **27**, 322–329 (1988).
- Guo, H. *et al.* Stimulation of matrix metalloproteinase production by recombinant extracellular matrix metalloproteinase inducer from transfected Chinese hamster ovary cells. *J. Biol. Chem.* **272**, 24–27 (1997).
- Anstee, D. J. The nature and abundance of human red cell surface glycoproteins. *J. Immunogenet.* **17**, 219–225 (1990).
- Theron, M., Hesketh, R. L., Subramanian, S. & Rayner, J. C. An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by *Plasmodium falciparum* parasites. *Cytometry A* **77A**, 1067–1074 (2010).
- Neafsey, D. E. *et al.* Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* **9**, R171 (2008).
- Bei, A. K., Brugnara, C. & Duraisingh, M. T. *In vitro* genetic analysis of an erythrocyte determinant of malaria infection. *J. Infect. Dis.* **202**, 1722–1727 (2010).
- Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Jallow, M. *et al.* Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genet.* **41**, 657–665 (2009).
- Teo, Y. Y., Small, K. S. & Kwiatkowski, D. P. Methodological challenges of genome-wide association analysis in Africa. *Nature Rev. Genet.* **11**, 149–160 (2010).
- Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the Ok^{a+} blood donors. We thank V. Horejsi for monoclonal antibodies and D. Ahr for technical assistance. This work was supported by the Wellcome Trust grant numbers 077108 (G.J.W.) and 089084 (J.C.R.) and National Institutes of Health R01AI057919 (M.T.D.). A.K.B. is supported by a Center for Disease Control grant R36 CK000119-01 and an Epidemiology of Infectious Disease and Biodefense Training Grant 2T32 AI007535-12.

Author Contributions C.C. compiled the erythrocyte protein library and identified the PfRh5–BSG interaction. L.Y.B. led the *P. falciparum* functional validation, with support from M.T. S.J.B. performed the biochemical and biophysical characterization of the interaction. A.K.B. performed the lentiviral knockdown and parasite invasion experiments under the direction of M.T.D. M.U. provided the Ok^{a+} blood samples and matching controls. O.N. and S.M. supervised the collection and culturing of field strains. D.P.K. performed genetic analysis on the BSG and PfRh5 loci. G.J.W. and J.C.R. conceived and supervised the project, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.W. (gw2@sanger.ac.uk) and J.C.R. (jr9@sanger.ac.uk).

METHODS

Recombinant protein production. Proteins selected for expression included all type I, type II, GPI (glycophosphatidylinositol)-linked receptors and secreted proteins. Some multipass transmembrane proteins were also included where there was an extracellular N terminus preceded by a signal peptide (Supplementary Table 1). Individual domains of human BSG were produced by identifying domain boundaries using the structure of the BSG extracellular region^{28,29} and amplifying these regions using primers with flanking NotI and AscI restriction enzyme sites to facilitate cloning. The carboxy-terminal amino acid sequence of the BSG-d0+1 and BSG-d1 constructs was HGPP. BSG-d2 was cloned into the same vector as *PfRh5* to add an exogenous signal peptide required for protein secretion and encompassed the sequence between PPRV.. and ..RSHL. Glycosylation sites were removed in BSG by mutating codons encoding all three asparagines in glycosylation motifs to aspartic acid. To remove N-linked glycans from soluble recombinant BSG, 500 units of PNGase F (New England Biolabs) were added to 10 µl of a spent tissue culture supernatant and incubated for 15 min at 37 °C. Sialic acid residues were removed by adding 1.6 milli-units of *Vibrio cholerae* neuraminidase (Sigma) to 10 µl of a spent tissue culture supernatant and incubated for 15 min at 37 °C.

Interaction screening by AVExis. For the AVExis assay, bait and prey protein preparations were normalized to activities that have been previously shown to detect transient interactions (monomeric half-lives less than 0.1 s) with a low false positive rate¹. Biotinylated baits dialysed against HBS were immobilised in the wells of a streptavidin-coated 96-well microtitre plate (NUNC). Normalized preys were added, incubated for 2 h at room temperature, washed three times in HBS/0.1% Tween-20, and once in HBS. 125 µg ml⁻¹ of nitrocefin was added, and absorbance values measured at 485 nm on a Pherastar plus (BMG laboratories). Controls were essentially as described¹² and included the Cd4d3+4 tag alone as a negative control bait and a biotinylated anti-Cd4 (anti-prey) antibody as a prey capture positive control. A positive control interaction consisting of the rat Cd200 bait detected using the rat Cd200R prey used at the threshold level and both 1:10 and 1:100 dilutions was included on each plate. The negative (–) and positive (+) control interactions shown in Fig. 1a are the rat Cd200R prey used at the screening threshold probed against the Cd4d3+4 (–) or rat Cd200 (+) baits.

***P. falciparum* culture, characterization and invasion assays.** All *P. falciparum* parasite strains were routinely cultured in human O+ erythrocytes at 5% haematocrit in complete medium (RPMI-1640 containing 10% human serum), under an atmosphere of 1% O₂, 3% CO₂ and 96% N₂. To confirm their identity, laboratory-adapted strains were genotyped by PCR within polymorphic regions of the *msp1* and *msp2* genes³⁰. Parasite cultures were synchronized in early stages with 5% (w/v) D-sorbitol (Sigma). Use of erythrocytes from human donors for *P. falciparum* culture was approved by the NHS Cambridgeshire 4 Research Ethics Committee. Ok^{a+} blood was obtained from donors in Japan with informed consent, and shipped on ice. For each sample, a control Ok^{a+} sample was collected at the same time under identical conditions. All experiments were performed within 72 h of collection.

Invasion assays were carried out in round-bottom 96-well plates, with a culture volume of 100 µl per well at a haematocrit of 2%. Parasites in trophozoite stage were mixed with pentamerized BSG-S–Cd4d3+4–COMP–His ectodomains or with anti-BSG monoclonal antibodies and incubated in the plates for 24 h at 37 °C inside a static incubator culture chamber (VWR), gassed with 1% O₂, 3% CO₂ and 96% N₂. At the end of the incubation period, red blood cells (RBC) were collected and parasitized RBC (pRBC) were stained with 2 µM Hoechst 33342 (Invitrogen), as described previously²¹. Invasion assays using Ok^{a+} blood and control Ok^{a+} blood were carried out following the two-colour flow cytometric assay described in ref. 21. Briefly, Ok^{a+} blood and control Ok^{a+} blood were labelled with 10 µM DDAO-SE (Invitrogen). RBC were resuspended to 2% haematocrit, mixed with pRBC (ring stage) and incubated in 96-well plates for 48 h as described above. At the end of the incubation period, RBC were harvested and pRBC were stained with 2 µM Hoechst 33342. Standard blood smear microscopy was performed to determine parasitaemia. Briefly, a small aliquot of the culture was smeared on a glass slide, fixed with 100% methanol and stained with Field's Stain (Pro-Lab Diagnostics). Parasitaemia was determined by counting the number of parasitized pRBC per 2,000 total RBC examined by oil immersion with a Leica DME microscope (Leica Microsystems). All parasitaemia represented was the average of three replicates. Lentiviral transductions of HSCs was performed as previously described²³. Lentiviral-delivered shRNA sequences were BSG; TRC clone ID (TRCN000006736) hairpin sequence: 5'-GAAGTCGTGACAAACACA TCAACTCGAGTTGATGTGTTCTGACGACTTC-3', pLKO scrambled control; (Addgene plasmid 1864) hairpin sequence: 5'-CCTAAGGTTAAGTCGCCCTC GCTCGAGCGAGGGCGCACTTAACCTTAGG-3'; loop region indicated in bold.

Detailed Standard Operating Procedures for all invasion assays are available at <http://www.sanger.ac.uk/research/>.

Flow cytometry. Stained samples were examined with a 355-nm ultraviolet laser (20 mW) and a 633-nm red laser (17 mW) on a BD LSRII flow cytometer (BD Biosciences). Hoechst 33342 (Invitrogen) was excited using the ultraviolet laser and detected with a 450/50 filter, whereas DDAO-SE (Invitrogen) was excited using the red laser and detected with a 660/20 filter. BD FACS Diva (BD Biosciences) was used to collect 100,000 events for each sample. FSC and SSC voltages of 423 and 198, respectively, and a threshold of 2,000 on FSC were applied to gate the erythrocyte population. The data collected were further analysed with FlowJo (Tree Star). All experiments were carried out in triplicate. GraphPad Prism (GraphPad Software) was used to plot the generated parasitaemia data.

PfRh5 cloning and sequencing. Total RNA was extracted from 3D7 and FCR3 schizonts using the QIAamp RNA Blood Mini Kit (Qiagen). Isolated RNA was treated with TURBO DNase (Ambion) and reverse transcribed to cDNA using the High-Capacity cDNA Archive Kit (Applied Biosystems) following the manufacturer's instructions. A 10 µl aliquot of cDNA was used as a template in a standard PCR reaction, using the primers Rh5-F (5'-ATGATAAGAATAAAAAAAAAA ATTAATTTTGACCATT-3') and Rh5-R (5'-TCATTGTGTAAGTGGTTTAT TTTTATATATGTTG-3'). Amplified fragments were subcloned into pCR2.1-TOPO, using the TOPO TA Cloning Kit (Invitrogen) and three clones from each strain were sequenced and analysed.

Antibodies. Antibodies were obtained from the following suppliers: anti-rat Cd4d3+4 (OX68) (AbD Serotec), anti-CD59 (AbD Serotec), mouse IgG₁ control (Abcam). Anti-BSG monoclonal antibodies used were 8J251 (Lifespan Biosciences), MEM-M6/1 (Abcam) and TRA-1-85 (R&D systems). MEM-M6/6 was provided as an ascitic fluid and was a gift of V. Horejsi; the antibody was purified using a HiTrap protein G column (GE Healthcare) as described³¹ and exchanged into RPMI.

Surface plasmon resonance. Surface plasmon resonance studies were performed using a Biacore T100 instrument. Briefly, biotinylated bait proteins were captured on a streptavidin-coated sensor chip (Biacore, GE Healthcare). Approximately 150 response units (RU) of the negative control bait (biotinylated rat Cd4d3+4) were immobilised in the flow cell used as a reference and approximate molar equivalents of the query protein immobilised in other flow cells. Purified analyte proteins were separated by gel filtration just before use in SPR experiments to remove small amounts of protein aggregates which are known to influence kinetic binding measurements³². Increasing concentrations of purified proteins were injected at high flow rates (100 µl min⁻¹) to minimise rebinding effects for kinetic studies or at 10 µl min⁻¹ for equilibrium analysis. Although essentially all the bound PfRh5 dissociated during the wash out phase (see Fig. 1c), the surface was 'regenerated' with a pulse of 2 M NaCl at the end of each cycle. Duplicate injections of the same concentration in each experiment were superimposable, demonstrating no loss of activity after regenerating the surface. Both kinetic and equilibrium binding data were analysed in the manufacturer's Biacore T100 evaluation software (Biacore). Equilibrium binding measurements were taken once equilibrium had been reached using reference-subtracted sensorgrams. Both the kinetic and equilibrium binding studies involving BSG-S and variants were performed three times using independent protein preparations of both PfRh5 and the BSG proteins, and once for BSG-L and its variants. All experiments were performed at 37 °C.

Enzyme-linked immunosorbent assay (ELISA). Biotinylated ectodomains were immobilized on streptavidin-coated plates (Nunc) for 1 h before incubation for 90 min with 10 µg ml⁻¹ primary antibody. The plates were washed in HBS/0.1% Tween-20 (HBST) before incubation with an appropriate secondary antibody conjugated to alkaline phosphatase (Sigma). Plates were washed three times in HBST and once in HBS before adding 100 µl *p*-nitrophenyl phosphate (Sigma 104 alkaline phosphatase substrate) at 1 mg ml⁻¹. Optical density measurements were taken at 405 nm on a Pherastar plus (BMG laboratories). The whole procedure was performed at room temperature.

28. Schlegel, J. *et al.* Solution characterization of the extracellular region of CD147 and its interaction with its enzyme ligand cyclophilin A. *J. Mol. Biol.* **391**, 518–535 (2009).
29. Yu, X. L. *et al.* Crystal structure of HAb18G/CD147: implications for immunoglobulin superfamily homophilic adhesion. *J. Biol. Chem.* **283**, 18056–18065 (2008).
30. Snounou, G. & Beck, H. P. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitol. Today* **14**, 462–467 (1998).
31. Crosnier, C., Staudt, N. & Wright, G. J. A rapid and scalable method for selecting recombinant mouse monoclonal antibodies. *BMC Biol.* **8**, 76 (2010).
32. van der Merwe, P. A. & Barclay, A. N. Analysis of cell-adhesion molecule interactions using surface plasmon resonance. *Curr. Opin. Immunol.* **8**, 257–261 (1996).

Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*

Cécile Crosnier^{1*}, Leyla Y. Bustamante^{2*}, S. Josefin Bartholdson^{1*}, Amy K. Bei³, Michel Theron², Makoto Uchikawa⁴, Souleymane Mboup⁵, Omar Ndir⁵, Dominic P. Kwiatkowski^{2,6}, Manoj T. Duraisingh³, Julian C. Rayner² & Gavin J. Wright^{1,2}

Erythrocyte invasion by *Plasmodium falciparum* is central to the pathogenesis of malaria. Invasion requires a series of extracellular recognition events between erythrocyte receptors and ligands on the merozoite, the invasive form of the parasite. None of the few known receptor–ligand interactions involved^{1–4} are required in all parasite strains, indicating that the parasite is able to access multiple redundant invasion pathways⁵. Here, we show that we have identified a receptor–ligand pair that is essential for erythrocyte invasion in all tested *P. falciparum* strains. By systematically screening a library of erythrocyte proteins, we have found that the Ok blood group antigen, basigin, is a receptor for PfRh5, a parasite ligand that is essential for blood stage growth⁶. Erythrocyte invasion was potently inhibited by soluble basigin or by basigin knockdown, and invasion could be completely blocked using low concentrations of anti-basigin antibodies; importantly, these effects were observed across all laboratory-adapted and field strains tested. Furthermore, Ok^{a-} erythrocytes, which express a basigin variant that has a weaker binding affinity for PfRh5, had reduced invasion efficiencies. Our discovery of a cross-strain dependency on a single extracellular receptor–ligand pair for erythrocyte invasion by *P. falciparum* provides a focus for new anti-malarial therapies.

Among the many *P. falciparum* merozoite proteins that are believed to have a role in erythrocyte invasion, most attention has focused on two major parasite protein families: the EBAs and Rh^s. Although erythrocyte receptors have been identified for some of them (members of the glycophorin family are receptors for three EBAs^{1–3} and complement receptor 1 (CD35) has recently been identified as a receptor for PfRh4, ref. 4) none of these receptor–ligand pairs are essential in all parasite strains tested. PfRh5 is unique amongst the EBAs and Rh^s because it cannot be deleted in any *P. falciparum* strain and is therefore apparently essential for parasite growth in blood stage culture^{5,6}. Both native and recombinant PfRh5 have been previously shown to bind erythrocytes through an unknown glycosylated receptor that is resistant to chymotrypsin, trypsin and neuraminidase treatment^{6,8,9}.

To identify an erythrocyte receptor for PfRh5, we used a systematic screening approach by first compiling a library of abundant cell surface and secreted proteins expressed by human erythrocytes based on published proteomics data¹⁰. Proteins for which the entire ectodomain was expected to be expressed as a soluble recombinant protein were selected (Supplementary Table 1), and expressed by mammalian cells (Supplementary Fig. 1). The 40 proteins within the erythrocyte ectodomain protein library were then systematically screened using the AVEXIS assay (avidity-based extracellular interaction screen)¹¹ for interactions with a recombinant PfRh5 protein, also produced by mammalian cells. The AVEXIS assay is designed to detect direct low-affinity protein interactions between ectodomain fragments expressed as either biotin-tagged baits or highly avid pentameric β -lactamase-tagged

preys^{12,13}. The PfRh5 prey interacted with a single erythrocyte receptor bait (Fig. 1a, top panel) corresponding to the Ok blood group antigen, basigin (BSG, also known as CD147, EMMPRIN and M6, ref. 14). The same single interaction was identified in the reciprocal bait–prey orientation (Fig. 1a, lower panel).

BSG is a member of the immunoglobulin superfamily (IgSF) and has been implicated in many biological functions including embryo implantation, spermatogenesis¹⁵ and retinal development¹⁶. BSG exists in both long (three IgSF domains, BSG-L) and short (two IgSF domains, BSG-S) splice isoforms (Fig. 1b) and although BSG-L was used in the screen, BSG-S is thought to be the major isoform expressed on erythrocytes. Binding experiments using domain deletions established that PfRh5 could interact with BSG-S and this required both domains because neither of the two BSG-S IgSF domains were individually able to bind PfRh5 (Fig. 1b and Supplementary Fig. 2). We showed that PfRh5 interacted directly with BSG-S and BSG-L using purified proteins and surface plasmon resonance (SPR). Both kinetic (Fig. 1c) and equilibrium (Supplementary Fig. 3) binding parameters for the interaction were derived using a 1:1 binding model and were in excellent agreement (Supplementary Table 2). These parameters are typical of extracellular protein interactions measured using this technique¹⁷. Removal of glycans from BSG either by mutating all predicted glycosylation motifs or by enzymatic treatment did not affect PfRh5 binding (Supplementary Fig. 4), indicating that the PfRh5 binding site is solely located in the BSG protein core. BSG is also known to be resistant to trypsin and chymotrypsin treatment¹⁸, consistent with previous PfRh5–erythrocyte binding studies^{6,8,9}.

To determine whether the PfRh5–BSG interaction was required for invasion, we added purified pentamerized soluble BSG-S into invasion assays to specifically compete with the membrane-bound receptor. We found that BSG-S strongly inhibited invasion in a dose-dependent manner relative to controls which included each of the two non-binding BSG-S IgSF domains added individually (Fig. 2a). Strong inhibition was also observed across multiple strains (Fig. 2b) or when soluble BSG-L was added (Supplementary Fig. 5), although this was slightly weaker for the 3D7 strain. Soluble forms of BSG consisting of the extracellular regions are known to have biological effects such as upregulation of matrix metalloproteases¹⁹. To rule out an indirect effect of exogenous BSG on invasion, we added to invasion assays two independent purified anti-BSG monoclonal antibodies (MEM-M6/6 and TRA-1-85) which could both block the PfRh5–BSG interaction *in vitro* (data not shown). These high-affinity reagents gave a potent invasion blocking effect that was saturable at very low antibody concentrations (half-maximum inhibitory concentration $\approx 0.5 \mu\text{g ml}^{-1}$), consistent with binding and occluding a specific surface receptor of typical abundance ($\sim 10^4$ to 10^6 molecules per cell²⁰) (Fig. 2c). Pre-adsorption of the MEM-M6/6 antibody with soluble monomeric BSG specifically relieved the inhibition, ruling out any indirect effect of the antibody on non-BSG targets;

¹Cell Surface Signalling Laboratory, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ²Malaria Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.

³Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁴Tokyo Red Cross Blood Center, Tokyo 135-8639, Japan. ⁵Laboratory of Bacteriology and Virology, Le Dantec Hospital and Laboratory of Parasitology, Cheikh Anta Diop University, BP: 7325, Dakar, Senegal. ⁶Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.

*These authors contributed equally to this work.

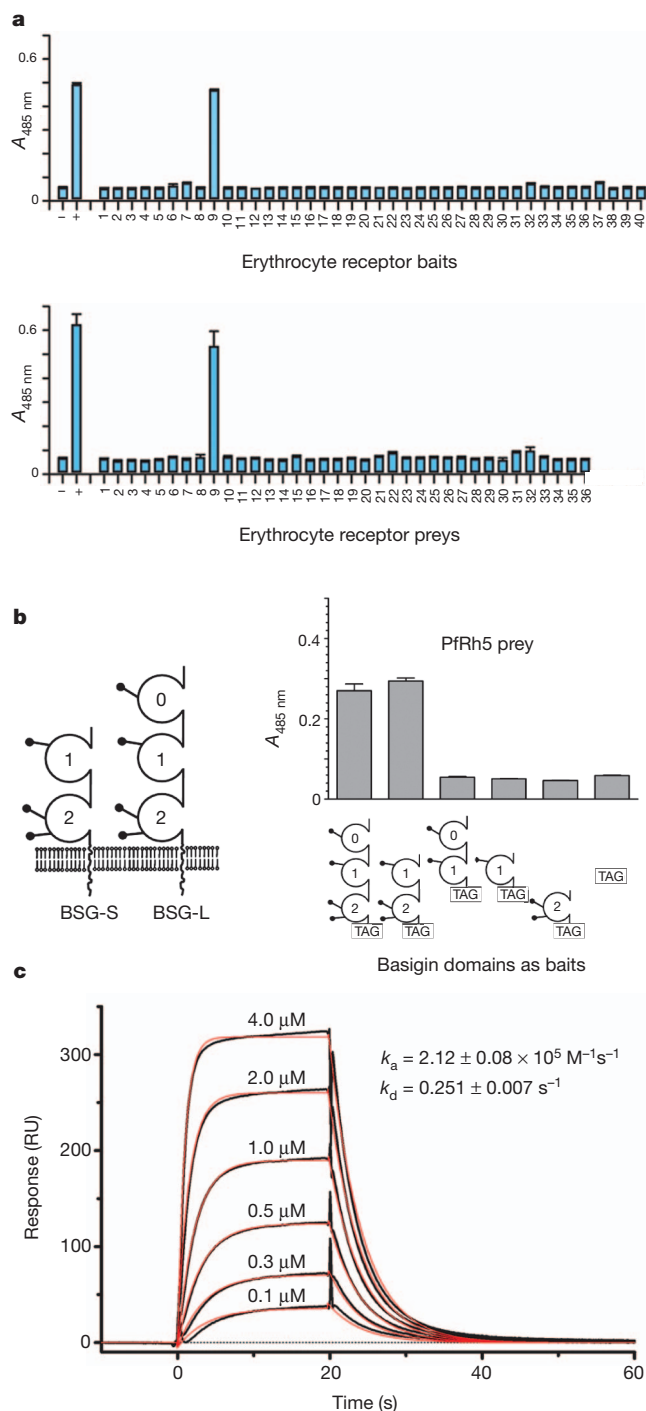


Figure 1 | BSG is an erythrocyte receptor for PfRh5. **a**, PfRh5 was screened as either a prey (top panel) or a bait (bottom panel) against an erythrocyte receptor protein library using AVEIXIS. BSG (protein 9) was identified as a receptor for PfRh5 in both bait–prey orientations. **b**, Domain structure of the BSG isoforms (left); lollipops represent potential N-linked glycosylation sites. BSG regions were expressed as baits and used to map the PfRh5 binding site to the two membrane-proximal domains. Bar charts show mean \pm s.e.m.; $n = 3$. **c**, Biophysical analysis of the PfRh5–BSG-S interaction using SPR. The indicated concentrations of purified PfRh5 were injected over immobilised BSG, and biophysical parameters derived from a 1:1 binding model (red line). RU, response units.

furthermore, MEM-M6/6 did not affect intra-erythrocytic *P. falciparum* development (Supplementary Fig. 6). Invasion was quantified using flow cytometry and a fluorescent DNA dye to stain parasites²¹. Using this assay, apparent invasion could not be eliminated, with efficiencies

reduced to a maximum of 80–90%, even at much higher concentrations of antibody (up to 1.5 mg ml^{-1} of MEM-M6/6, data not shown); however, direct observation of parasites using Giemsa-stained thin smears revealed that this residual staining in cytometry assays was due to extracellular parasites and debris in the culture. Using microscopy-based assays, we found that MEM-M6/6 concentrations of $10 \mu\text{g ml}^{-1}$ or more was sufficient to prevent all detectable invasion (Fig. 2d).

P. falciparum isolates can vary widely in their ability to invade erythrocytes treated with different receptor-modifying enzymes such as trypsin, chymotrypsin and neuraminidase, showing differential dependencies on erythrocyte receptors for invasion. To determine if BSG was a critical invasion receptor across *P. falciparum* lines that use different invasion pathways, we tested the ability of MEM-M6/6 to block erythrocyte invasion on nine culture-adapted strains representing seven different PfRh5 sequence variants (Supplementary Table 3). We observed that the invasion of all lines was potently inhibited by MEM-M6/6 (Fig. 2e). To show that the dependency on BSG was not an unusual feature of culture-adapted lines, we also tested six freshly-isolated *P. falciparum* strains from Senegal²² and again observed a potent inhibitory effect (Fig. 2f). Assays with the field isolates were carried out with unsynchronised parasites, decreasing the overall inhibitory effect because not all parasites had reinvasion over the course of the assay. All six Senegal isolates, however, were inhibited by MEM-M6/6 to the same extent as an unsynchronised culture-adapted line, W2mef, tested at the same time. This demonstrated that freshly-isolated field strains have the same dependency on BSG as laboratory-adapted lines (Fig. 2f).

To confirm independently the essentiality of BSG as a *P. falciparum* invasion receptor, we used a genetic approach by differentiating erythrocytes from haematopoietic stem cells transduced with lentiviruses containing either a short hairpin RNA targeting *BSG* or a scrambled shRNA control (pLKO). BSG-targeted erythrocytes showed a reproducible knockdown to approximately 50 to 60% of cell surface BSG levels relative to the pLKO control (Fig. 2g) and expressed markers indicative of complete erythrocyte maturation (Supplementary Fig. 7). The invasion of both the 3D7 and W2mef *P. falciparum* strains into BSG-knockdown erythrocytes was significantly reduced compared to the control (18% versus 94% for 3D7 and 14% versus 103% for W2mef, Fig. 2h). By contrast, previous knockdown of GYPA, the major surface sialoglycoprotein, significantly inhibited the W2mef but not the 3D7 strain²³. The inhibition of erythrocyte invasion by multiple *P. falciparum* strains using soluble BSG, anti-BSG monoclonal antibodies, or knockdown of BSG surface expression suggests that BSG is a critical host receptor for *P. falciparum* invasion.

Malaria is thought to have been a strong selective pressure in human evolutionary history and given the apparently essential roles of PfRh5 and BSG in *P. falciparum* invasion we sought to determine if any human populations contained genetic variants in *BSG* that might affect PfRh5 binding and invasion. Five nonsynonymous single nucleotide polymorphisms (SNPs) have been described within the BSG-S IgSF domains (Supplementary Table 4 and Fig. 3a). These variants were expressed and the biophysical PfRh5 binding parameters determined using SPR. Equilibrium measurements showed that two variants had lower binding affinity compared to the BSG reference sequence: L90P and E92K (Fig. 3b and Supplementary Table 2). L90P did not interact with PfRh5 and binding profiles of several anti-BSG monoclonal antibodies indicated local misfolding of the membrane-distal IgSF domain (Supplementary Fig. 8). No verification or population frequency data for this SNP are currently available, preventing further biological interpretation of this variant. E92K had a twofold lower affinity for PfRh5 (Fig. 3b) and a comparative kinetic analysis demonstrated that this was due to both a slower association and a faster dissociation rate (Fig. 3c and Supplementary Table 2). The E92 residue is solvent-exposed and located within the loop connecting

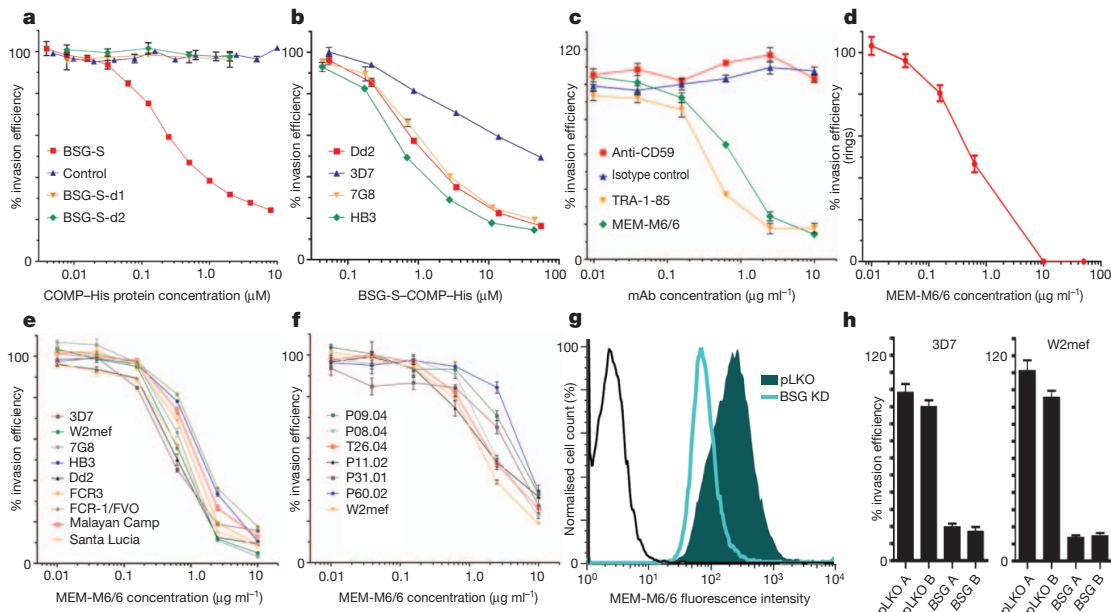


Figure 2 | Soluble BSG, anti-BSG antibodies and BSG knockdown potently block erythrocyte invasion. **a**, Erythrocyte invasion was inhibited by purified pentamerized BSG-S-Cd4d3+4-COMP-His ectodomains but not by the two non-binding BSG-S domains added individually or Cd4d3+4-COMP-His (control); strain Dd2. **b**, Cross-strain inhibition of invasion using pentamerized BSG-S. **c**, Anti-BSG monoclonal antibodies, TRA-1-85 and MEM-M6/6, potently inhibited invasion of erythrocytes; strain 3D7. **d**, MEM-M6/6 concentrations $\geq 10 \mu\text{g ml}^{-1}$ prevented all detectable invasion by microscopic

observation of cultures; strain 3D7. **e**, **f**, MEM-M6/6 inhibited invasion of synchronised *P. falciparum* culture-adapted lines (**e**) and unsynchronised field isolates (**f**). **g**, Cell-surface BSG is reduced in erythrocytes differentiated from haematopoietic stem cells transduced with lentiviruses containing shRNA targeting BSG (light blue line) relative to a control virus (pLKO, shaded); black line represents secondary antibody alone. **h**, 3D7 and W2mef invasion was inhibited in BSG knockdown erythrocytes. A and B are replicates. Invasion efficiencies are mean \pm s.e.m., $n = 3$.

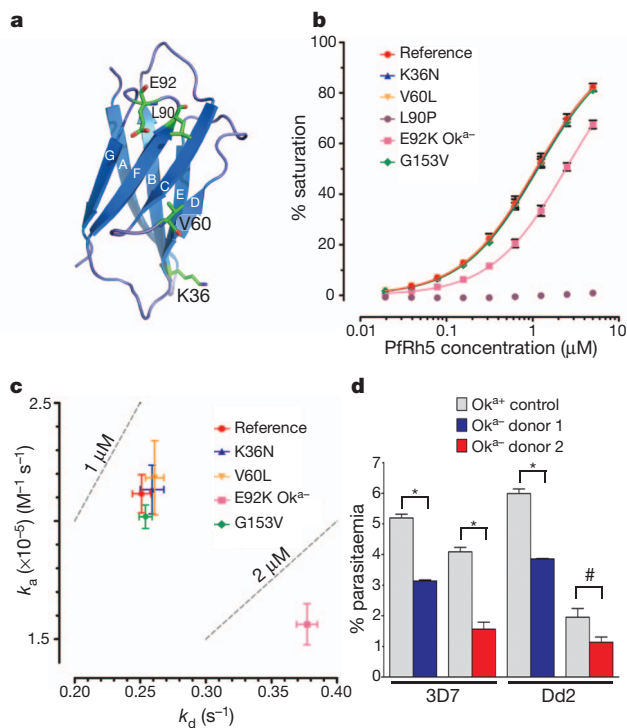


Figure 3 | The Ok^{a-} BSG variant has reduced binding affinity for PfRh5 and Ok^{a-} erythrocytes have reduced merozoite invasion frequencies. **a**, Schematic of the membrane distal IgSF domain of BSG-S showing the location of naturally-occurring variants. **b**, Equilibrium binding isotherms of PfRh5 binding to BSG-S variants. **c**, Association (k_a) and dissociation (k_d) rate constants of PfRh5 binding to BSG-S and variants. Means \pm s.e.m.; $n = 3$. **d**, Invasion of 3D7 and Dd2 strains in Ok^{a-} blood cells are reduced relative to the Ok^{a+} control. Mean \pm s.e.m., $n = 3$; * $P \leq 0.0003$; # $P = 0.0349$, unpaired one-tailed *t* test. A repeat is shown in Supplementary Fig. 9.

the F-G β -strands close to the glycan-free GFC β -sheet, consistent with a possible PfRh5-binding interface (Fig. 3a). E92K is the variant responsible for the Ok^{a-} blood group, which has been described in eight Japanese families¹⁴. Ok^{a-} erythrocytes from two unrelated donors showed reduced invasion with both 3D7 and Dd2 *P. falciparum* strains relative to Ok^{a+} controls (Fig. 3d and Supplementary Fig. 9), correlating with the reduced affinity of the Ok^{a-} variant for PfRh5. The extreme rarity and restriction of the Ok^{a-} blood group to Japanese individuals suggest that this specific allele has not had a major role in conferring resistance to malaria. It is possible that other BSG polymorphisms, as yet unknown, have evolved in some malaria-exposed populations as a mechanism of resistance to *P. falciparum*. The search for functional polymorphisms of BSG needs to go beyond gene coding regions as the results of our knockdown experiments indicate that expression levels of BSG at the erythrocyte surface influence the ability of the parasite to invade. The Duffy variant, which confers resistance to *P. vivax*, is also a non-coding regulatory polymorphism that suppresses expression of the invasion receptor by erythrocytes. Our ability to address this problem is currently limited by the lack of data on genome variation among the many different ethnic groups that are exposed to *P. falciparum* malaria, but will be greatly enhanced by the 1000 Genomes Project, MalariaGEN and other genetic studies that are now in progress in Africa and other malaria-endemic regions of the world^{24–26}. Inter-population comparisons of haplotype length and frequency provide a potentially powerful way of addressing this problem²⁷, and there is preliminary evidence that a region of chromosome 19 encompassing BSG and several neighbouring genes has undergone recent positive selection in West Africa, but a considerable amount of further work is needed to determine whether this is causally related to the role of BSG as a malaria invasion receptor (MalariaGEN consortium, unpublished data).

In summary, we have applied a systematic protein interaction screening approach (AVEXIS) to identify BSG as an erythrocyte receptor for PfRh5. Importantly, we were able to prevent all detectable erythrocyte invasion by every *P. falciparum* strain that we tested using only modest concentrations of anti-BSG antibodies. These

observations, coupled with the inability to delete *PfRh5*⁶, lead us to conclude that the interaction between BSG and PfRh5 is essential for parasite entry, and may perform a fundamentally different function to the other EBA and Rh proteins, which are involved in redundant, partially overlapping invasion pathways. The dependence on a single receptor–ligand pair across many *P. falciparum* strains may provide new possibilities for therapeutic intervention.

METHODS SUMMARY

Recombinant protein production and interaction screening. Protein production, purification, AVExis assays and SPR were performed essentially as described¹¹ except the type II proteins which were expressed with an amino-terminal C4d3+4-biotin tag and a mouse antibody signal peptide. PfRh5 was expressed as above except that a non-endogenous signal peptide was added and the threonines in potential N-linked glycan sequons were mutated to alanine to prevent inappropriate glycosylation. All constructs were chemically synthesized and codon optimised for mammalian expression (GeneArt AG). Purified pentameric proteins used in invasion assays were made by replacing the β -lactamase reporter in the prey plasmid with a hexa-his tag, purified and buffer exchanged into RPMI before use. BSG variants were produced by site directed mutagenesis.

***P. falciparum* culture, lentiviral transduction and invasion assays.** *P. falciparum* parasite strains were routinely cultured in human O+ erythrocytes at 5% haematocrit. Use of erythrocytes from human donors for *P. falciparum* culture was approved by the NHS Cambridgeshire 4 Research Ethics Committee. Ok^{a+} and control Ok^{a+} blood was obtained from donors in Japan with informed consent, shipped on ice and experiments performed within 72 h. Invasion assays were carried out as described previously²¹ using the two-colour assay for the Ok^{a+} experiment. Lentiviral transduction of HSCs was performed as previously described²³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 March; accepted 3 October 2011.

Published online 9 November 2011; corrected 21 December 2011 (see full-text HTML version for details).

- Maier, A. G. *et al.* *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nature Med.* **9**, 87–92 (2003).
- Mayer, D. C. *et al.* Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proc. Natl Acad. Sci. USA* **106**, 5348–5352 (2009).
- Sim, B. K. *et al.* Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* **264**, 1941–1944 (1994).
- Tham, W. H. *et al.* Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proc. Natl Acad. Sci. USA* **107**, 17327–17332 (2010).
- Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–766 (2006).
- Baum, J. *et al.* Reticulocyte-binding protein homologue 5 – an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *Int. J. Parasitol.* **39**, 371–380 (2009).
- Iyer, J. *et al.* Invasion of host cells by malaria parasites: a tale of two protein families. *Mol. Microbiol.* **65**, 231–249 (2007).
- Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4**, 40–51 (2008).
- Rodriguez, M. *et al.* PfRH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE* **3**, e3300 (2008).
- Pasini, E. M. *et al.* In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **108**, 791–801 (2006).
- Bushell, K. M. *et al.* Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* **18**, 622–630 (2008).
- Martin, S. *et al.* Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol. Cell. Proteomics* **9**, 2654–2665 (2010).
- Söllner, C. & Wright, G. J. A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biol.* **10**, R99 (2009).
- Spring, F. A. *et al.* The Ok^a blood group antigen is a marker for the M6 leukocyte activation antigen, the human homolog of OX-47 antigen, basigin and neurothelin, an immunoglobulin superfamily molecule that is widely expressed in human cells and tissues. *Eur. J. Immunol.* **27**, 891–897 (1997).
- Igakura, T. *et al.* A null mutation in basigin, an immunoglobulin superfamily member, indicates its important roles in peri-implantation development and spermatogenesis. *Dev. Biol.* **194**, 152–165 (1998).
- Fadool, J. M. & Linser, P. J. 5A11 antigen is a cell recognition molecule which is involved in neuronal-glial interactions in avian neural retina. *Dev. Dyn.* **196**, 252–262 (1993).
- Wright, G. J. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. Biosyst.* **5**, 1405–1412 (2009).
- Williams, B. P. *et al.* Biochemical and genetic analysis of the Ok^a blood group antigen. *Immunogenetics* **27**, 322–329 (1988).
- Guo, H. *et al.* Stimulation of matrix metalloproteinase production by recombinant extracellular matrix metalloproteinase inducer from transfected Chinese hamster ovary cells. *J. Biol. Chem.* **272**, 24–27 (1997).
- Anstee, D. J. The nature and abundance of human red cell surface glycoproteins. *J. Immunogenet.* **17**, 219–225 (1990).
- Theron, M., Hesketh, R. L., Subramanian, S. & Rayner, J. C. An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by *Plasmodium falciparum* parasites. *Cytometry A* **77A**, 1067–1074 (2010).
- Neafsey, D. E. *et al.* Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* **9**, R171 (2008).
- Bei, A. K., Brugnara, C. & Duraisingh, M. T. *In vitro* genetic analysis of an erythrocyte determinant of malaria infection. *J. Infect. Dis.* **202**, 1722–1727 (2010).
- Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Jallow, M. *et al.* Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genet.* **41**, 657–665 (2009).
- Teo, Y. Y., Small, K. S. & Kwiatkowski, D. P. Methodological challenges of genome-wide association analysis in Africa. *Nature Rev. Genet.* **11**, 149–160 (2010).
- Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the Ok^{a+} blood donors. We thank V. Horejsi for monoclonal antibodies and D. Ahr for technical assistance. This work was supported by the Wellcome Trust grant numbers 077108 (G.J.W.) and 089084 (J.C.R.) and National Institutes of Health R01AI057919 (M.T.D.). A.K.B. is supported by a Center for Disease Control grant R36 CK000119-01 and an Epidemiology of Infectious Disease and Biodefense Training Grant 2T32 AI007535-12.

Author Contributions C.C. compiled the erythrocyte protein library and identified the PfRh5–BSG interaction. L.Y.B. led the *P. falciparum* functional validation, with support from M.T. S.J.B. performed the biochemical and biophysical characterization of the interaction. A.K.B. performed the lentiviral knockdown and parasite invasion experiments under the direction of M.T.D. M.U. provided the Ok^{a+} blood samples and matching controls. O.N. and S.M. supervised the collection and culturing of field strains. D.P.K. performed genetic analysis on the BSG and PfRh5 loci. G.J.W. and J.C.R. conceived and supervised the project, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.W. (gw2@sanger.ac.uk) and J.C.R. (jr9@sanger.ac.uk).

METHODS

Recombinant protein production. Proteins selected for expression included all type I, type II, GPI (glycophosphatidylinositol)-linked receptors and secreted proteins. Some multipass transmembrane proteins were also included where there was an extracellular N terminus preceded by a signal peptide (Supplementary Table 1). Individual domains of human BSG were produced by identifying domain boundaries using the structure of the BSG extracellular region^{28,29} and amplifying these regions using primers with flanking NotI and AscI restriction enzyme sites to facilitate cloning. The carboxy-terminal amino acid sequence of the BSG-d0+1 and BSG-d1 constructs was HGPP. BSG-d2 was cloned into the same vector as *PfRh5* to add an exogenous signal peptide required for protein secretion and encompassed the sequence between PPRV.. and ..RSHL. Glycosylation sites were removed in BSG by mutating codons encoding all three asparagines in glycosylation motifs to aspartic acid. To remove N-linked glycans from soluble recombinant BSG, 500 units of PNGase F (New England Biolabs) were added to 10 µl of a spent tissue culture supernatant and incubated for 15 min at 37 °C. Sialic acid residues were removed by adding 1.6 milli-units of *Vibrio cholerae* neuraminidase (Sigma) to 10 µl of a spent tissue culture supernatant and incubated for 15 min at 37 °C.

Interaction screening by AVExis. For the AVExis assay, bait and prey protein preparations were normalized to activities that have been previously shown to detect transient interactions (monomeric half-lives less than 0.1 s) with a low false positive rate¹. Biotinylated baits dialysed against HBS were immobilised in the wells of a streptavidin-coated 96-well microtitre plate (NUNC). Normalized preys were added, incubated for 2 h at room temperature, washed three times in HBS/0.1% Tween-20, and once in HBS. 125 µg ml⁻¹ of nitrocefin was added, and absorbance values measured at 485 nm on a Pherastar plus (BMG laboratories). Controls were essentially as described¹² and included the Cd4d3+4 tag alone as a negative control bait and a biotinylated anti-Cd4 (anti-prey) antibody as a prey capture positive control. A positive control interaction consisting of the rat Cd200 bait detected using the rat Cd200R prey used at the threshold level and both 1:10 and 1:100 dilutions was included on each plate. The negative (–) and positive (+) control interactions shown in Fig. 1a are the rat Cd200R prey used at the screening threshold probed against the Cd4d3+4 (–) or rat Cd200 (+) baits.

***P. falciparum* culture, characterization and invasion assays.** All *P. falciparum* parasite strains were routinely cultured in human O+ erythrocytes at 5% haematocrit in complete medium (RPMI-1640 containing 10% human serum), under an atmosphere of 1% O₂, 3% CO₂ and 96% N₂. To confirm their identity, laboratory-adapted strains were genotyped by PCR within polymorphic regions of the *msp1* and *msp2* genes³⁰. Parasite cultures were synchronized in early stages with 5% (w/v) D-sorbitol (Sigma). Use of erythrocytes from human donors for *P. falciparum* culture was approved by the NHS Cambridgeshire 4 Research Ethics Committee. Ok^{a+} blood was obtained from donors in Japan with informed consent, and shipped on ice. For each sample, a control Ok^{a+} sample was collected at the same time under identical conditions. All experiments were performed within 72 h of collection.

Invasion assays were carried out in round-bottom 96-well plates, with a culture volume of 100 µl per well at a haematocrit of 2%. Parasites in trophozoite stage were mixed with pentamerized BSG-S–Cd4d3+4–COMP–His ectodomains or with anti-BSG monoclonal antibodies and incubated in the plates for 24 h at 37 °C inside a static incubator culture chamber (VWR), gassed with 1% O₂, 3% CO₂ and 96% N₂. At the end of the incubation period, red blood cells (RBC) were collected and parasitized RBC (pRBC) were stained with 2 µM Hoechst 33342 (Invitrogen), as described previously²¹. Invasion assays using Ok^{a+} blood and control Ok^{a+} blood were carried out following the two-colour flow cytometric assay described in ref. 21. Briefly, Ok^{a+} blood and control Ok^{a+} blood were labelled with 10 µM DDAO-SE (Invitrogen). RBC were resuspended to 2% haematocrit, mixed with pRBC (ring stage) and incubated in 96-well plates for 48 h as described above. At the end of the incubation period, RBC were harvested and pRBC were stained with 2 µM Hoechst 33342. Standard blood smear microscopy was performed to determine parasitaemia. Briefly, a small aliquot of the culture was smeared on a glass slide, fixed with 100% methanol and stained with Field's Stain (Pro-Lab Diagnostics). Parasitaemia was determined by counting the number of parasitized pRBC per 2,000 total RBC examined by oil immersion with a Leica DME microscope (Leica Microsystems). All parasitaemia represented was the average of three replicates. Lentiviral transductions of HSCs was performed as previously described²³. Lentiviral-delivered shRNA sequences were BSG; TRC clone ID (TRCN000006736) hairpin sequence: 5'-GAAGTCGTGACAAACACA TCAACTCGAGTTGATGTGTTCTGACGACTTC-3', pLKO scrambled control; (Addgene plasmid 1864) hairpin sequence: 5'-CCTAAGGTTAAGTCGCCCTC GCTCGAGCGAGGGCGCACTTAACCTTAGG-3'; loop region indicated in bold.

Detailed Standard Operating Procedures for all invasion assays are available at <http://www.sanger.ac.uk/research/>.

Flow cytometry. Stained samples were examined with a 355-nm ultraviolet laser (20 mW) and a 633-nm red laser (17 mW) on a BD LSRII flow cytometer (BD Biosciences). Hoechst 33342 (Invitrogen) was excited using the ultraviolet laser and detected with a 450/50 filter, whereas DDAO-SE (Invitrogen) was excited using the red laser and detected with a 660/20 filter. BD FACS Diva (BD Biosciences) was used to collect 100,000 events for each sample. FSC and SSC voltages of 423 and 198, respectively, and a threshold of 2,000 on FSC were applied to gate the erythrocyte population. The data collected were further analysed with FlowJo (Tree Star). All experiments were carried out in triplicate. GraphPad Prism (GraphPad Software) was used to plot the generated parasitaemia data.

PfRh5 cloning and sequencing. Total RNA was extracted from 3D7 and FCR3 schizonts using the QIAamp RNA Blood Mini Kit (Qiagen). Isolated RNA was treated with TURBO DNase (Ambion) and reverse transcribed to cDNA using the High-Capacity cDNA Archive Kit (Applied Biosystems) following the manufacturer's instructions. A 10 µl aliquot of cDNA was used as a template in a standard PCR reaction, using the primers Rh5-F (5'-ATGATAAGAATAAAAAAAAAA ATTAATTTTGACCATT-3') and Rh5-R (5'-TCATTGTGTAAGTGGTTTAT TTTTATATATGTTG-3'). Amplified fragments were subcloned into pCR2.1-TOPO, using the TOPO TA Cloning Kit (Invitrogen) and three clones from each strain were sequenced and analysed.

Antibodies. Antibodies were obtained from the following suppliers: anti-rat Cd4d3+4 (OX68) (AbD Serotec), anti-CD59 (AbD Serotec), mouse IgG₁ control (Abcam). Anti-BSG monoclonal antibodies used were 8J251 (Lifespan Biosciences), MEM-M6/1 (Abcam) and TRA-1-85 (R&D systems). MEM-M6/6 was provided as an ascitic fluid and was a gift of V. Horejsi; the antibody was purified using a HiTrap protein G column (GE Healthcare) as described³¹ and exchanged into RPMI.

Surface plasmon resonance. Surface plasmon resonance studies were performed using a Biacore T100 instrument. Briefly, biotinylated bait proteins were captured on a streptavidin-coated sensor chip (Biacore, GE Healthcare). Approximately 150 response units (RU) of the negative control bait (biotinylated rat Cd4d3+4) were immobilised in the flow cell used as a reference and approximate molar equivalents of the query protein immobilised in other flow cells. Purified analyte proteins were separated by gel filtration just before use in SPR experiments to remove small amounts of protein aggregates which are known to influence kinetic binding measurements³². Increasing concentrations of purified proteins were injected at high flow rates (100 µl min⁻¹) to minimise rebinding effects for kinetic studies or at 10 µl min⁻¹ for equilibrium analysis. Although essentially all the bound PfRh5 dissociated during the wash out phase (see Fig. 1c), the surface was 'regenerated' with a pulse of 2 M NaCl at the end of each cycle. Duplicate injections of the same concentration in each experiment were superimposable, demonstrating no loss of activity after regenerating the surface. Both kinetic and equilibrium binding data were analysed in the manufacturer's Biacore T100 evaluation software (Biacore). Equilibrium binding measurements were taken once equilibrium had been reached using reference-subtracted sensorgrams. Both the kinetic and equilibrium binding studies involving BSG-S and variants were performed three times using independent protein preparations of both PfRh5 and the BSG proteins, and once for BSG-L and its variants. All experiments were performed at 37 °C.

Enzyme-linked immunosorbent assay (ELISA). Biotinylated ectodomains were immobilized on streptavidin-coated plates (Nunc) for 1 h before incubation for 90 min with 10 µg ml⁻¹ primary antibody. The plates were washed in HBS/0.1% Tween-20 (HBST) before incubation with an appropriate secondary antibody conjugated to alkaline phosphatase (Sigma). Plates were washed three times in HBST and once in HBS before adding 100 µl *p*-nitrophenyl phosphate (Sigma 104 alkaline phosphatase substrate) at 1 mg ml⁻¹. Optical density measurements were taken at 405 nm on a Pherastar plus (BMG laboratories). The whole procedure was performed at room temperature.

28. Schlegel, J. *et al.* Solution characterization of the extracellular region of CD147 and its interaction with its enzyme ligand cyclophilin A. *J. Mol. Biol.* **391**, 518–535 (2009).
29. Yu, X. L. *et al.* Crystal structure of HAb18G/CD147: implications for immunoglobulin superfamily homophilic adhesion. *J. Biol. Chem.* **283**, 18056–18065 (2008).
30. Snounou, G. & Beck, H. P. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. *Parasitol. Today* **14**, 462–467 (1998).
31. Crosnier, C., Staudt, N. & Wright, G. J. A rapid and scalable method for selecting recombinant mouse monoclonal antibodies. *BMC Biol.* **8**, 76 (2010).
32. van der Merwe, P. A. & Barclay, A. N. Analysis of cell-adhesion molecule interactions using surface plasmon resonance. *Curr. Opin. Immunol.* **8**, 257–261 (1996).

Response to self antigen imprints regulatory memory in tissues

Michael D. Rosenblum^{1*}, Iris K. Gratz^{2*}, Jonathan S. Paw^{1,2}, Karen Lee³, Ann Marshak-Rothstein⁴ & Abul K. Abbas²

Immune homeostasis in tissues is achieved through a delicate balance between pathogenic T-cell responses directed at tissue-specific antigens and the ability of the tissue to inhibit these responses. The mechanisms by which tissues and the immune system communicate to establish and maintain immune homeostasis are currently unknown. Clinical evidence suggests that chronic or repeated exposure to self antigen within tissues leads to an attenuation of pathological autoimmune responses, possibly as a means to mitigate inflammatory damage and preserve function. Many human organ-specific autoimmune diseases are characterized by the initial presentation of the disease being the most severe, with subsequent flares being of lesser severity and duration¹. In fact, these diseases often spontaneously resolve, despite persistent tissue autoantigen expression². In the practice of antigen-specific immunotherapy, allergens or self antigens are repeatedly injected in the skin, with a diminution of the inflammatory response occurring after each successive exposure³. Although these findings indicate that tissues acquire the ability to attenuate autoimmune reactions upon repeated responses to antigens, the mechanism by which this occurs is unknown. Here we show that upon expression of self antigen in a peripheral tissue, thymus-derived regulatory T cells (T_{reg} cells) become activated, proliferate and differentiate into more potent suppressors, which mediate resolution of organ-specific autoimmunity in mice. After resolution of the inflammatory response, activated T_{reg} cells are maintained in the target tissue and are primed to attenuate subsequent autoimmune reactions when antigen is re-expressed. Thus, T_{reg} cells function to confer 'regulatory memory' to the target tissue. These findings provide a framework for understanding how T_{reg} cells respond when exposed to self antigen in peripheral tissues and offer mechanistic insight into how tissues regulate autoimmunity.

We hypothesized that exposure of immune cells to self antigen in peripheral tissues induces stable regulatory mechanisms that limit autoimmune injury. To test this and to define the nature of these control mechanisms, we created a novel mouse model of inducible tissue-specific self-antigen expression. We crossed transgenic mice expressing a membrane-bound form of ovalbumin (Ova) under the control of a tetracycline response element to transgenic mice expressing the tetracycline transactivator protein under the control of the keratin 5 (K5, also called Krt5) promoter (Fig. 1a)⁴. In the resultant K5/TGO double transgenic mice (see Methods for definition of TGO), Ova expression in the skin is tightly controlled, as adoptively transferred Ova-specific DO11.10 (DO11) $CD4^{+}$ T cells become activated and proliferate in skin-draining lymph nodes (SDLNs) of recipient mice only after treatment with doxycycline (Fig. 1b). In addition, Ova messenger RNA is detected in epidermal cell suspensions only after induction with doxycycline (Supplementary Fig. 1).

To define functional Ova expression in the thymus, K5/TGO mice were crossed with the DO11.10 T-cell receptor (TCR)-transgenic

strain⁵. Deletion of $CD4^{+}$ DO11 T cells or increased production of $Foxp3^{+}CD4^{+}$ DO11 T_{reg} cells in K5/TGO/DO11 mice is a sensitive indicator of thymic Ova expression. K5/TGO/DO11 triple transgenic mice as well as TGO/DO11 double transgenic mice have modest deletion of $CD4^{+}$ DO11 cells (Fig. 1c) and a pronounced increase in DO11 T_{reg} cells in both the thymus and SDLNs (Fig. 1c, d). Between 30–40% of antigen-specific $CD4^{+}$ DO11 T cells in the SDLNs of K5/TGO/DO11 (and TGO/DO11) mice co-express CD25 and $Foxp3$ (Fig. 1d). These results demonstrate that Ova is constitutively expressed in the thymus, independent of doxycycline treatment and dependent only on the presence of the TGO transgene. Thus, K5/TGO/DO11 mice represent a unique model in which antigen is continuously expressed in the thymus and tightly controlled in the periphery, mimicking the pattern of tissue-specific self-antigen expression in mice and humans^{6,7}.

Despite the presence of a large percentage of Ova-specific T_{reg} cells, induction of cutaneous Ova expression in K5/TGO/DO11 mice results in a pronounced inflammatory dermatitis (Fig. 2a). Disease peaks at

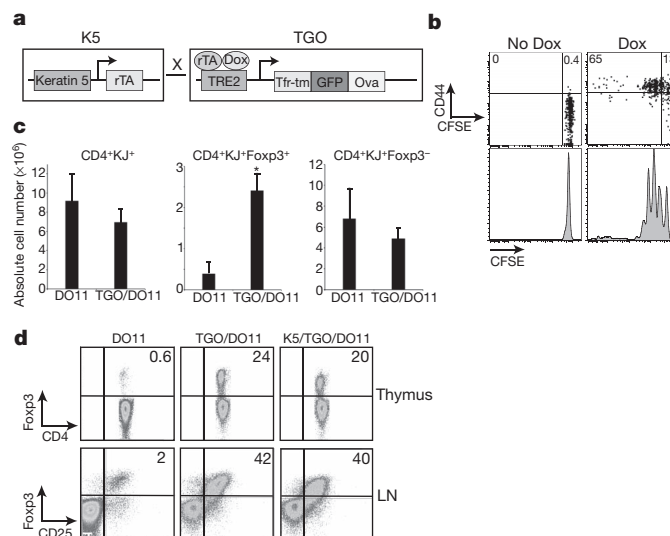


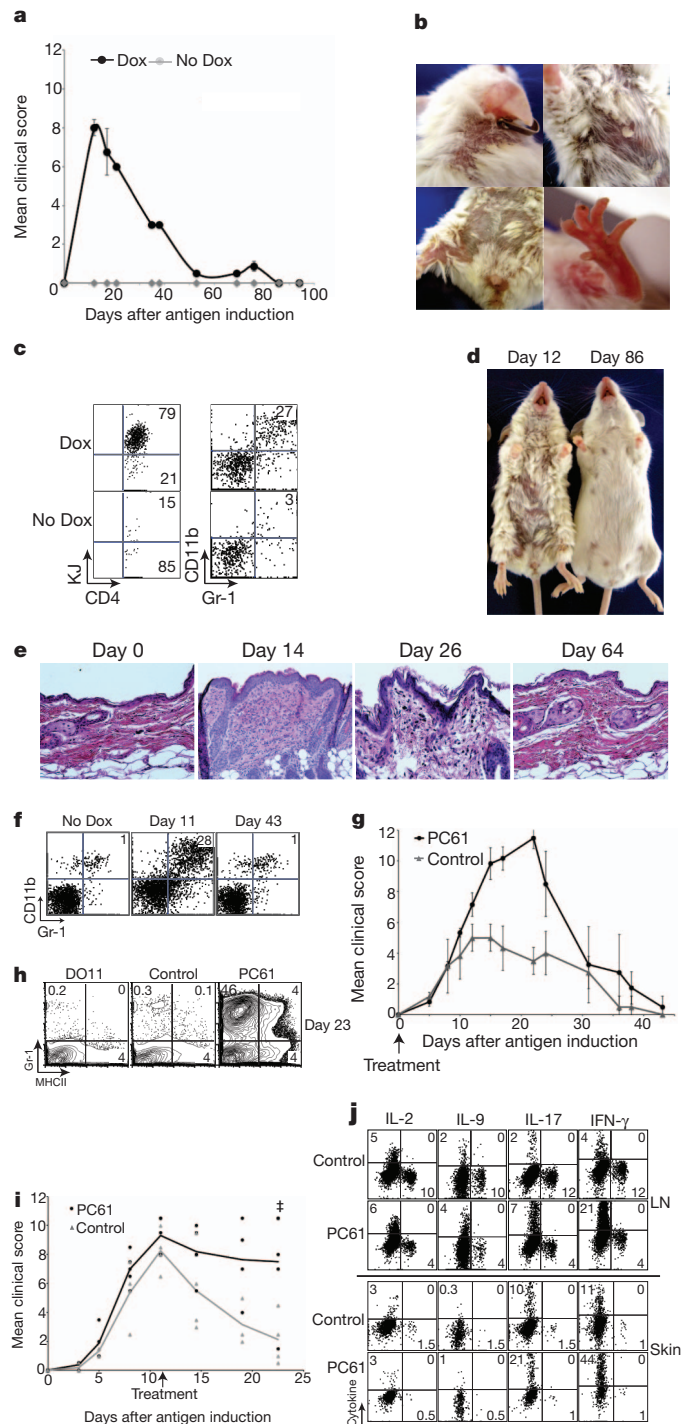
Figure 1 | Characterization of K5/TGO/DO11 mice. **a**, Construct for double transgenic mice expressing ovalbumin (Ova) driven by the cytokeratin 5 (K5) promoter in a tetracycline-inducible fashion. **b**, Lymph node cells from DO11 TCR-transgenic mice were labelled with CFSE and injected into K5/TGO mice. Recipient mice were fed doxycycline chow and DO11 cell proliferation (CFSE dilution) and CD44 expression was measured 3 days later by flow cytometry. **c**, SDLN cell numbers from DO11 and TGO/DO11 mice in the absence of doxycycline treatment. **d**, Thymus and SDLN cells from DO11, TGO/DO11 and K5/TGO/DO11 mice in the absence of doxycycline treatment. Thymocytes are gated on $CD4^{+}KJ^{+}CD8^{-}$ cells. KJ, KJ1-26 monoclonal antibody. Lymph node cells are gated on $CD4^{+}KJ^{+}$ cells. * $P < 0.05$ (t -test). Error bars represent s.d. Results are representative of 3 replicate experiments with $n = 3$ –4 mice per group.

¹Department of Dermatology, University of California San Francisco, San Francisco, California 94115, USA. ²Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA. ³Department of Pediatrics, Columbia University Medical Center, New York, New York 10032, USA. ⁴Department of Medicine, Rheumatology Division, University of Massachusetts, Worcester, Massachusetts 01655, USA.

*These authors contributed equally to this work.

10–14 days after antigen induction and is characterized by marked erythema, scaling and alopecia (Fig. 2b). The skin infiltrate at the height of disease is composed primarily of antigen-specific DO11 T cells and GR-1⁺CD11b⁺ myeloid cells (Fig. 2c). Ova-specific skin-infiltrating CD4⁺ T cells produce interferon (IFN)- γ and interleukin (IL)-17 (Fig. 2j).

Interestingly, skin inflammation spontaneously resolves. Twenty to 30 days after antigen induction, mice begin to show signs of clinical and histological improvement, and by 40 to 60 days, there is complete resolution of cutaneous inflammation (Fig. 2d–f). Disease resolves despite continued doxycycline treatment, with ongoing Ova expression in the skin and continuous thymic output of Ova-specific T cells (Supplementary Fig. 2).



Given the large percentage of antigen-specific T_{reg} cells present in the SDLNs of K5/TGO/DO11 mice, it is surprising that cutaneous inflammation results when antigen is induced. We were further intrigued by the spontaneous resolving nature of the autoimmune response, as it suggests that an immune regulatory mechanism(s) is initiated in an attempt to minimize tissue damage. To determine if T_{reg} cells have a role in resolving skin inflammation, we depleted these cells before antigen induction and at the height of disease. Treatment of K5/TGO/DO11 mice with CD25-depleting antibody (PC61) before antigen induction resulted in pronounced disease exacerbation with delayed resolution of skin inflammation (Fig. 2g, h). In some experiments, depletion of T_{reg} cells resulted in dehydration, weight loss and death, most likely from severely disrupted skin barrier function (Supplementary Fig. 3). Depletion of T_{reg} cells at the height of disease (that is, 10 days after antigen induction) resulted in delayed kinetics of resolution of skin inflammation, with some mice having to be euthanized secondary to non-resolving skin disease (Fig. 2i). Higher percentages of both IFN- γ and IL-17-secreting DO11 cells were observed in the skin and SDLN after T_{reg} cell depletion (Fig. 2j). Anti-CD25 antibody was used in these studies because CD25 is constitutively expressed on T_{reg} cells and antibody treatment can be accurately titrated to achieve optimal deletion of these cells (Supplementary Fig. 4). A caveat of this approach is that other CD25-expressing cells (which may be either tolerance-promoting or pathogenic) might also be deleted. The finding that anti-CD25 antibody treatment inhibits or prevents disease resolution strongly suggests that T_{reg} cells are the major targets of the antibody and have an essential role in resolving skin disease in our model; however, depletion of other CD25-expressing regulatory populations cannot be definitively ruled out.

The emergence of disease despite the presence of T_{reg} cells and the T_{reg} cell-dependent suppression of disease after antigen induction suggested that expression of tissue antigen activates T_{reg} cells and enhances their suppressive function. To test this, we induced Ova expression in K5/TGO/DO11 mice and characterized T_{reg} cells in both the skin and SDLNs. Foxp3⁺ DO11 cells robustly expanded in SDLNs upon antigen induction (Fig. 3a). Expansion peaked at 12–14 days, when these cells were increased ~15-fold over baseline numbers. In contrast, the peak expansion of Foxp3⁻ DO11 effector cells occurred earlier and these cells only expanded ~4-fold over baseline (Fig. 3a). At the peak of expansion, more than 50% of Foxp3⁺ DO11 cells had markedly reduced CD25 on the cell surface (Fig. 3b). Loss of CD25 was transient, associated only with the proliferative burst, as Foxp3⁺ DO11 cells in SDLNs during the resolution phase of disease (>20 days) express CD25 (Fig. 3b). Loss of CD25 expression on *in vivo* proliferating T_{reg} cells (with preservation of suppressive capacity)

Figure 2 | K5/TGO/DO11 mice develop autoimmune skin disease that resolves spontaneously. **a**, Mean clinical scores of K5/TGO/DO11 mice left untreated or fed doxycycline chow. **b**, Skin lesions in a representative K5/TGO/DO11 mouse at the height of clinical disease. **c**, Flow cytometry of skin-infiltrating cells in K5/TGO/DO11 mice treated with doxycycline for 12 days. **d**, Representative mice from the height of disease (day 12) and after disease resolution (day 86). **e**, Skin histology of K5/TGO/DO11 mice after beginning doxycycline. **f**, Flow cytometry of skin-infiltrating cells in K5/TGO/DO11 mice at 11 and 43 days after beginning doxycycline. **g**, Disease scores of K5/TGO/DO11 mice treated with anti-CD25 monoclonal antibody (PC61) or control (PBS or isotype control antibody) at 10 days and 3 days before beginning doxycycline. **h**, Flow cytometry of skin-infiltrating cells in K5/TGO/DO11 mice untreated or treated with PC61 at 23 days after beginning doxycycline. **i**, Disease scores of K5/TGO/DO11 mice treated with PC61 or isotype control antibody at the height of disease. Results are pooled data from two replicate experiments with $n = 3-4$ mice per group. **j**, Intracellular cytokine stains of SDLN cells and skin-infiltrating cells from PC61- or isotype control-treated (before antigen induction) K5/TGO/DO11 mice at 11 days after beginning doxycycline treatment. Gated on CD4⁺KJ⁺ cells. All error bars represent s.d. Results are representative of 3 replicate experiments with $n = 3-4$ mice per group. Euthanization of two PC61-treated mice secondary to non-resolving skin disease is indicated by the ‡ symbol.

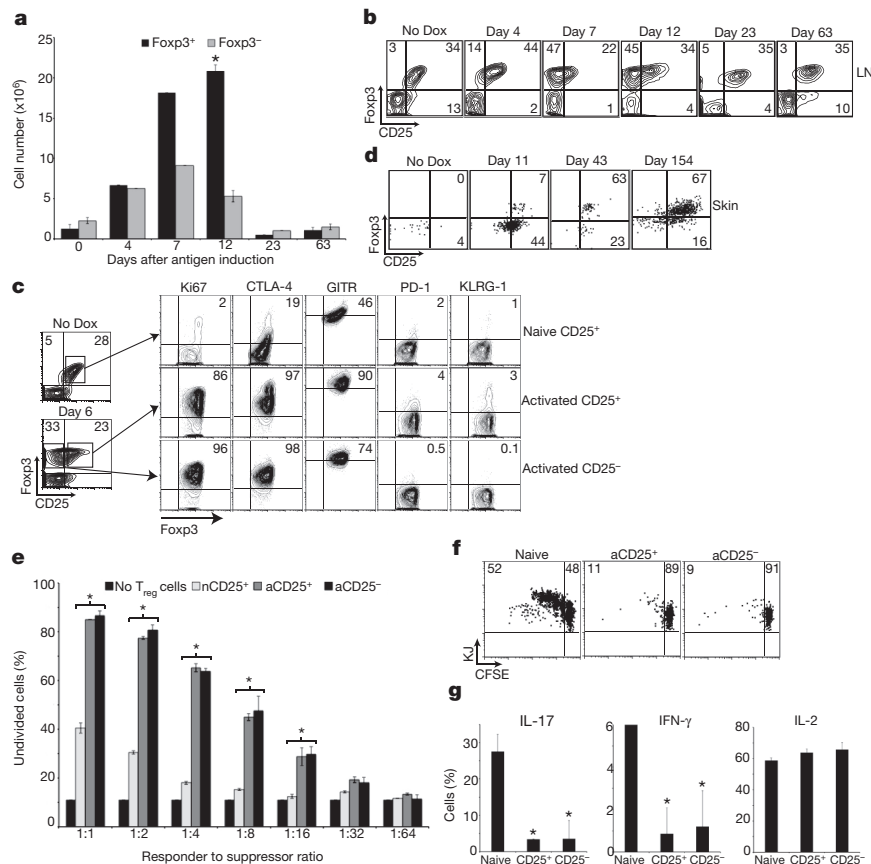


Figure 3 | T_{reg} cells are activated upon induction of peripheral antigen.

a, Quantification of CD4⁺KJ⁺Foxp3⁺ and CD4⁺KJ⁺Foxp3⁻ cell numbers from SDLNs of K5/TGO/DO11 mice after beginning doxycycline treatment. **b**, Flow cytometry of SDLN cells from K5/TGO/DO11 mice after beginning doxycycline treatment. Gated on CD4⁺KJ⁺ cells. **c**, Flow cytometry of SDLN cells from K5/TGO/DO11 mice left untreated (naive CD25⁺) or treated for 6 days with doxycycline (activated CD25⁺ or CD25⁻ populations). **d**, Flow cytometry of skin-infiltrating cells from K5/TGO/DO11 mice after beginning doxycycline. Gated on live CD4⁺KJ⁺ cells. **e**, *In vitro* suppression assay using sorted CD4⁺KJ⁺Foxp3⁺CD25⁺ cells from non-induced naive mice (nCD25⁺), CD4⁺KJ⁺Foxp3⁺CD25⁺ cells from doxycycline-treated mice

has been reported by others⁸. It is possible that proliferating T_{reg} cells reduce expression of CD25 in a negative feedback loop to limit proliferation and regulate T_{reg} cell numbers. SDLN T_{reg} cells in the resolution phase of disease have increased expression of Foxp3 and CD25 when compared to T_{reg} cells from non-induced mice (Fig. 3b and Supplementary Fig. 5). These results indicate that T_{reg} cells are activated upon exposure to peripheral antigen. To test this further, we analysed expression levels of various markers shown to be involved in T_{reg} cell function early after antigen induction. Six days after inducing Ova expression in the skin, >80% of T_{reg} cells have entered the cell cycle, as shown by Ki67 expression, and proliferating CD25⁺ and CD25⁻ T_{reg} cells express higher levels of CTLA-4 compared to T_{reg} cells from non-induced mice (Fig. 3c).

Expansion of T_{reg} cells in SDLNs upon antigen induction is associated with an accumulation of these cells in the skin (Fig. 3d). Before antigen induction, CD4⁺ DO11 cells are barely detectable in the skin. At the height of disease, most skin-infiltrating DO11 cells are Foxp3⁻. However, resolution of inflammation is associated with a preferential accumulation of Foxp3⁺ cells in the skin, with >60% of DO11 cells in the skin expressing Foxp3 in mice maintained on doxycycline that have resolved disease (that is, clinically and histologically normal appearing skin) (Fig. 3d).

We next tested whether T_{reg} cells acquire a more suppressive phenotype after exposure to tissue antigen. To do so, we crossed K5/TGO/

(activated CD25⁺, aCD25⁺), or CD4⁺KJ⁺Foxp3⁺CD25⁻ cells from doxycycline-treated mice (aCD25⁻) as suppressors and CFSE-labelled DO11 lymph node cells as responders. **f**, *In vivo* suppression assay using suppressors (as in part **e**) mixed 1:1 with naive CFSE-labelled Thy1.1⁺ DO11 lymph node cells and adoptively transferred into K5/TGO mice. Recipient mice were treated with doxycycline and DO11 cell proliferation and intracellular cytokine expression (**g**) were assayed 3 days later. Gated on Thy1.1⁺CD4⁺KJ⁺ cells. **P* < 0.05 (*t*-test). All error bars represent s.d. Results are representative of 3 replicate experiments with *n* = 3–4 mice per group except for *in vitro* and *in vivo* suppression assays, which are representative of 2 replicate experiments with *n* = 2–3 mice per group.

DO11 mice with Foxp3^{GFP} reporter mice⁹. K5/TGO/DO11/Foxp3^{GFP} mice were treated with doxycycline to induce antigen expression in the skin, and 6 days later, CD4⁺Foxp3⁺CD25⁺ and CD4⁺Foxp3⁺CD25⁻ DO11 cells were isolated from SDLNs and their suppressive capacity tested *in vitro* and *in vivo*. Both CD25⁺ and CD25⁻ T_{reg} cells from antigen-induced mice were more potent suppressors of effector T-cell proliferation *in vitro* when compared to T_{reg} cells isolated from non-induced mice (that is, peripheral antigen naive) at all effector to suppressor ratios examined (Fig. 3e). In addition, T_{reg} cells from antigen-induced mice were more potent suppressors of cutaneous autoimmune responses *in vivo*. Upon co-transfer with naive DO11 responder cells into K5/TGO double transgenic mice (and treatment of recipient mice with doxycycline to induce Ova expression in the skin), both CD25⁺ and CD25⁻ Foxp3⁺ T_{reg} cells isolated from antigen-induced K5/TGO/DO11/Foxp3^{GFP} mice were more potent at inhibiting T-cell proliferation when compared to Foxp3⁺ cells isolated from non-induced mice (Fig. 3f). These cells were also more potent inhibitors of IL-17 and IFN-γ production from effector T cells (Fig. 3g).

The mechanism of enhanced T_{reg} cell suppression after peripheral antigen induction may involve CTLA-4, which is expressed at markedly higher levels after activation of T_{reg} cells (Fig. 3c) and has been shown to be a critical mediator of T_{reg} function in suppressing tissue inflammation¹⁰. In addition, CD25⁺ and CD25⁻ T_{reg} cells had similarly high

levels of CTLA-4 expression and are equally suppressive *in vitro* and *in vivo*. We were unable to detect cytokine production (including IL-10, IL-17 and IFN- γ) from T_{reg} cells before or after antigen induction, suggesting that a cytokine-mediated mechanism of suppression is unlikely (Fig. 2j and data not shown).

Our results indicate that early after exposure to tissue autoantigen, T_{reg} cells are activated, proliferate, differentiate into more potent suppressors, and progressively accumulate in the target tissue. These results are consistent with recent findings that two major populations of Foxp3⁺ T_{reg} cells exist in human peripheral blood, which have been called resting T_{reg} cells and activated T_{reg} cells¹¹. Resting T_{reg} cells are recently derived from the thymus, express low levels of CTLA-4, are not actively cycling, and are inferior suppressors when compared to activated T_{reg} cells. In contrast, activated T_{reg} cells are constitutively cycling, express higher levels of Foxp3 and CTLA-4, and are more potent suppressors. Longitudinal studies showed ongoing conversion of resting T_{reg} cells to activated T_{reg} cells in a healthy human volunteer, with T-cell-receptor clonotypes initially detected in the resting T_{reg} cell subset found 18 months later in the activated T_{reg} cell subset. These results suggest that in normal healthy individuals, resting T_{reg} cells are continuously activated and proliferate into more potent activated T_{reg} cells. Taken together with results presented here, we speculate that exposure to tissue self antigen may drive this process as a mechanism to inhibit and/or suppress autoimmunity.

We hypothesized that once T_{reg} cells have been activated and accumulate in the skin, these cells are relatively stable and persist in the tissue. They would then be capable of suppressing subsequent autoimmune reactions when antigen is re-expressed. To test this, we induced antigen expression in K5/TGO/DO11 mice and upon resolution of skin disease, removed doxycycline for >30 days to effectively 'turn antigen off' in the skin. We then re-induced antigen expression and followed mice clinically. To confirm the absence of antigen expression upon cessation of doxycycline, we assessed activation and proliferation of adoptively transferred DO11 cells and activation of endogenous DO11 cells in K5/TGO/DO11 mice (as a surrogate marker of antigen expression) at various times after discontinuing doxycycline. In addition, we measured Ova mRNA levels in the skin. At 20 days after discontinuing doxycycline treatment there was no proliferation of adoptively transferred DO11 cells and by 29 days, CD44 and CD69 expression had returned to baseline levels on adoptively transferred and endogenous DO11 cells, respectively (Fig. 4a and Supplementary Fig. 6a). Expression of Ova mRNA was undetectable in all mice at 32 days after stopping doxycycline treatment (Supplementary Fig. 6b). Taken together, these data confirm that antigen expression is effectively 'turned off' in all mice by 30 days after cessation of doxycycline treatment. Analysis of DO11 T cells in the skin of K5/TGO/DO11 mice that had been off doxycycline for >30 days revealed a persistent Foxp3⁺ population comprising >50% of DO11 cells in the skin (Fig. 4b). The proportion of DO11 cells in the skin that were Foxp3⁺ was enriched relative to SDLNs, indicating a preferential accumulation of T_{reg} cells in the target tissue upon cessation of antigen expression (Fig. 4b). Furthermore, T_{reg} cells in the skin continued to express high levels of CTLA-4 (Fig. 4c). When compared to T_{reg} cells in the SDLN, cells that persisted in the skin (in the absence of antigen expression) expressed lower levels of CD25 and higher levels of CTLA-4, CD127 and KLRG1 (Fig. 4d). The persistence of activated T_{reg} cells in the skin indicated that there may be an attenuation of cutaneous inflammation upon antigen re-expression, as T_{reg} cells must be present in the skin to suppress autoimmunity¹². To test this, we re-started doxycycline treatment and followed mice clinically. Upon antigen re-expression, K5/TGO/DO11 mice developed skin disease phenotypically similar to the disease that developed upon initial antigen expression; however, the severity of skin disease was significantly reduced and it resolved with accelerated kinetics (Fig. 4e). Attenuated disease upon antigen re-exposure was not secondary to reduced numbers of DO11 cells, as these cells returned to baseline

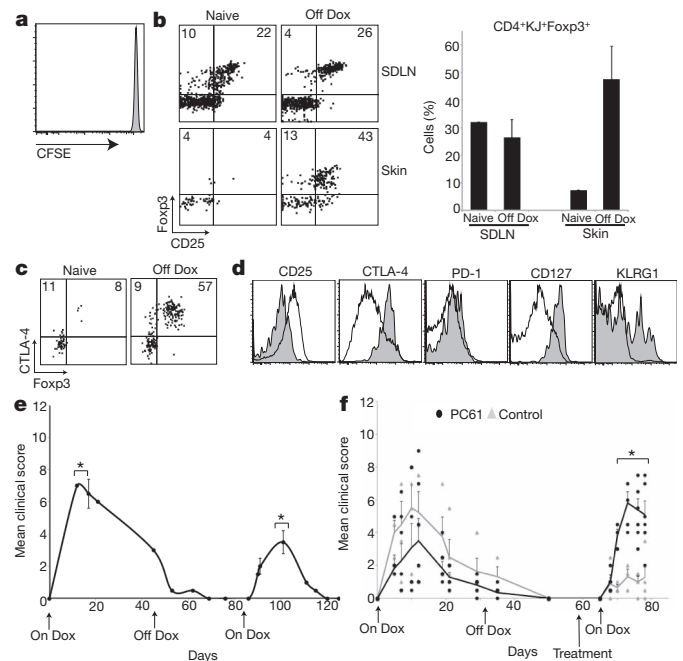


Figure 4 | Memory T_{reg} cells attenuate skin disease upon re-expression of tissue antigen. **a**, Flow cytometry of CFSE-labelled Thy1.1⁺DO11⁺ SDLN cells 3 days after adoptive transfer into K5/TGO/DO11 mice that had been on doxycycline for >30 days and off for 20 days. **b**, Flow cytometry and T_{reg} cell percentages of SDLN and skin-infiltrating CD4⁺ DO11 cells isolated from K5/TGO/DO11 mice that were on doxycycline for >30 days and off for >30 days. Gated on CD4⁺ KJ⁺ cells. **c**, Flow cytometry of CD4⁺ DO11 cells from the skin of K5/TGO/DO11 mice that have not been treated with doxycycline (naive) or on for >30 days and off doxycycline for >30 days. **d**, Phenotype of CD4⁺Foxp3⁺ DO11 cells from skin (shaded) and SDLN (unshaded) from K5/TGO/DO11 mice that were on doxycycline for >30 days and off for >40 days. **e**, Clinical skin disease upon re-starting doxycycline in K5/TGO/DO11 mice that had been on doxycycline for >30 days and off for >30 days. **f**, Clinical skin disease of K5/TGO/DO11 mice treated with either PC61 or isotype control antibody before re-starting doxycycline. In all mice, antigen was induced, skin disease developed, and disease had resolved before antibody treatment. All error bars represent standard error of samples within each group. Mean clinical scores from individual mice are shown. Results are representative of 3 replicate experiments with 3–4 mice per group except for **f**, which is combined data from 2 replicate experiments with 2–4 mice per group. **P* < 0.05 (*t*-test).

levels relatively early after cessation of doxycycline treatment (Supplementary Fig. 7). In addition, antigen-experienced DO11 T cells from K5/TGO/DO11 mice that had resolved disease and were either maintained on doxycycline or were taken off doxycycline after resolution readily induced skin disease when adoptively transferred into new antigen-expressing hosts (Supplementary Fig. 8). This suggests that cell-intrinsic anergy in responding T cells does not seem to have a major role in attenuating skin inflammation. To test whether reduced disease severity upon second antigen exposure was mediated by T_{reg} cells, we depleted these cells before re-inducing antigen expression. Depletion of T_{reg} cells resulted in a complete abrogation of the attenuated response observed upon re-expression of Ova in the skin (Fig. 4f).

Our results indicate that exposure to tissue autoantigens leads to the activation of self-reactive T_{reg} cells that were generated by self-antigen expression in the thymus. Activated T_{reg} cells persist in the target tissue and suppress autoimmune responses upon repeated or chronic encounters with tissue autoantigen. Thus, tissues that have undergone autoimmune inflammatory reactions develop a property we have termed regulatory memory, which serves to limit the severity of future such reactions. Memory T_{reg} cells have several characteristics of effector memory cells^{13,14}, including survival without antigen, residence in non-lymphoid tissues, phenotypic markers suggestive of prior activation, and enhanced functional activity. The life history of

T_{reg} cells is fundamentally similar to that of conventional T cells, passing through defined phases that include generation in the thymus, activation in the periphery leading to proliferation and differentiation into functionally more active cells, and survival as memory populations. Although our data support an obligatory role for T_{reg} cells in suppressing primary and memory responses in the skin, it is possible that these cells work together with other T_{reg} -cell-independent mechanisms to suppress tissue-specific autoimmunity. Identifying and harnessing the network of regulatory pathways that serve to limit tissue inflammation will undoubtedly offer new strategies for controlling autoimmune reactions and preserving end-organ function.

METHODS SUMMARY

Mouse model. TRE-TGO mice were crossed with K5-rtTA mice to generate mice with inducible epidermal expression of Ova. The TGO construct encodes a fusion protein linking the transferrin receptor transmembrane domain, green fluorescent protein (GFP) and amino acids 230–359 of chicken ovalbumin. To create K5/TGO/DO11 mice, K5-rtTA/TRE-TGO mice were crossed onto the DO11.10 TCR-transgenic background. In all experiments, mice were gender-matched and between 5 and 24 weeks of age. All mice were bred and maintained in a specific pathogen-free facility in accordance with the guidelines of the Laboratory Animal Resource Center of the University of California San Francisco.

Inflammatory skin disease model. To induce expression of the TGO transgene, K5/TGO/DO11 mice were maintained on 1 g kg^{-1} doxycycline chow. A 12-point clinical scoring scale was used to quantify skin disease. Scaling, alopecia, erythema and level of activity were each given a score from 0 to 3. Individual scores were summed and mean scores per group are displayed.

T_{reg} cell depletion. PC61 (anti-CD25 monoclonal antibody) or isotype control were injected intraperitoneally at 10 and 3 days before doxycycline treatment. For depletion at the height of disease, PC61 monoclonal antibody was injected on days 10 and 11 after starting K5/TGO/DO11 mice on doxycycline. For experiments in which CD25⁺ cells were deleted before antigen re-expression, K5/TGO/DO11 mice that had been off doxycycline for >30 days were treated with PC61 at 7 and 3 days before restarting doxycycline.

In vitro and in vivo suppression assays. K5/TGO/DO11 mice were crossed to Foxp3-GFP transgenic mice. K5/TGO/DO11/Foxp3-GFP mice were started on doxycycline and 6 days later, T_{reg} cells were isolated. For *in vitro* suppression assays, T_{reg} cells were cultured with CFSE-labelled SDLN cells isolated from DO11.10⁺/Rag2^{-/-}/Thy1.1⁺ mice and Ova peptide-pulsed bone-marrow-derived dendritic cells. For *in vivo* suppression assays, purified T_{reg} cells were mixed 1:1 with CFSE-labelled DO11 effector cells from DO11.10⁺/Rag2^{-/-}/Thy1.1⁺ mice. Cell mixtures were injected intravenously into K5/TGO mice and recipient mice were started on doxycycline.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 21 October 2011.

Published online 27 November 2011.

- James, W. D. *Andrews' Diseases of the Skin: Clinical Dermatology* (Saunders Elsevier, 2006).

- Lara-Corrales, I. & Pope, E. Autoimmune blistering diseases in children. *Semin. Cutan. Med. Surg.* **29**, 85–91 (2010).
- Sabatos-Peyton, C. A., Verhagen, J. & Wraith, D. C. Antigen-specific immunotherapy of autoimmune and allergic diseases. *Curr. Opin. Immunol.* **22**, 609–615 (2010).
- Diamond, I., Owolabi, T., Marco, M., Lam, C. & Glick, A. Conditional gene expression in the epidermis of transgenic mice using the tetracycline-regulated transactivators tTA and rTA linked to the keratin 5 promoter. *J. Invest. Dermatol.* **115**, 788–794 (2000).
- Murphy, K. M., Heimberger, A. B. & Loh, D. Y. Induction by antigen of intrathymic apoptosis of CD4⁺CD8⁺TCR^{lo} thymocytes *in vivo*. *Science* **250**, 1720–1723 (1990).
- Wada, N. *et al.* Aire-dependent thymic expression of desmoglein 3, the autoantigen in pemphigus vulgaris, and its role in T-cell tolerance. *J. Invest. Dermatol.* **131**, 410–417 (2011).
- Mouquet, H. *et al.* Expression of pemphigus-autoantigen desmoglein 1 in human thymus. *Tissue Antigens* **71**, 464–470 (2008).
- Gavin, M. A., Clarke, S. R., Negrou, E., Gallegos, A. & Rudensky, A. Homeostasis and anergy of CD4⁺CD25⁺ suppressor T cells *in vivo*. *Nature Immunol.* **3**, 33–41 (2002).
- Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* **22**, 329–341 (2005).
- Wing, K. *et al.* CTLA-4 control over Foxp3⁺ regulatory T cell function. *Science* **322**, 271–275 (2008).
- Miyara, M. *et al.* Functional delineation and differentiation dynamics of human CD4⁺ T cells expressing the FoxP3 transcription factor. *Immunity* **30**, 899–911 (2009).
- Dudda, J. C., Perdue, N., Bachtanian, E. & Campbell, D. J. Foxp3⁺ regulatory T cells maintain immune homeostasis in the skin. *J. Exp. Med.* **205**, 1559–1565 (2008).
- Kurtulus, S., Tripathi, P., Opferman, J. T. & Hildeman, D. A. Contracting the “mus cells”—does down-sizing suit us for diving into the memory pool? *Immunol. Rev.* **236**, 54–67 (2010).
- Akbar, A. N., Vukmanovic-Stejic, M., Taams, L. S. & Macallan, D. C. The dynamic co-evolution of memory and regulatory CD4⁺ T cells in the periphery. *Nature Rev. Immunol.* **7**, 231–237 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Benetiz for assistance with animal husbandry, S. Isakson for genotyping, S.-w. Jiang and M. Lee for cell sorting, and K. Ravid and G. Martin for derivation of TRE-TGO transgenic mice. We thank S. Ziegler, Benaroya Research Institute, for transgenic mice. M.D.R. is supported by a Dermatology Foundation Career Development Award and the UCSF Department of Dermatology. This work was partially funded through NIH grants P01 AI35297, R01 AI73656 and U19 AI56388 (to A.K.A.); NIH grant AR055634 to (A.M.-R.); and the Scleroderma Research Foundation (A.M.-R.). I.K.G. is supported by an Erwin Schrodinger Fellowship from the Austrian Science Fund (FWF), J2997-B13.

Author Contributions M.D.R. and I.K.G. contributed equally to this work and designed the studies, performed the experiments and analysed the data. M.D.R. and A.K.A. wrote the manuscript. J.S.P. collected and analysed data as well as helped with mouse husbandry. K.L. engineered and derived the TRE-TGO mice in the laboratory of A.M.-R. A.K.A. oversaw all study design and data analysis. A.M.-R. was involved in study design and data analysis. All authors discussed results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.K.A. (Abul.Abbas@ucsf.edu).

METHODS

Mouse model. K5/rTA mice were generated as described⁴. The TGO construct encodes for a fusion protein linking the transferrin receptor transmembrane domain (Tfr-tm), GFP and amino acids 230–359 of chicken ovalbumin (Ova)¹⁵. TGO was cloned upstream of the tetracycline response element (TRE2) and transgenic mice with stable incorporation of the TRE-TGO construct were generated. GFP is expressed at levels too low to detect in these mice. TRE-TGO mice were crossed with K5-rTA mice to generate mice with inducible epidermal expression of Ova. To create K5/TGO/DO11 mice, K5-rTA/TRE-TGO mice were crossed onto the DO11.10 TCR transgenic background, which expresses Ova-specific CD4⁺ T cells⁵. KJ1-26 monoclonal antibody (KJ) specifically recognizes the DO11 TCR. All mice were bred and maintained in a specific pathogen-free facility in accordance with the guidelines of the Laboratory Animal Resource Center of the University of California San Francisco.

T_{reg} cell depletion. Anti-CD25 monoclonal antibody (PC61) preferentially depletes T_{reg} cells¹⁶, especially in the BALB/c strain¹⁷. PC61 or isotype control (UCSF Monoclonal Antibody Core) were injected (0.5 mg per mouse) intraperitoneally at 10 and 3 days before doxycycline treatment. PC61 treatment of TGO/DO11 mice results in a > 90% reduction of CD4⁺CD25⁺Foxp3⁺ DO11 cells at 7 days after injection (Supplementary Fig. 4). For depletion of CD25⁺ cells at the height of disease, PC61 monoclonal antibody (0.5 mg per mouse) was injected on days 10 and 11 after starting K5/TGO/DO11 mice on doxycycline chow. For experiments in which CD25⁺ cells were deleted before antigen re-expression, K5/TGO/DO11 mice that had been off doxycycline for >30 days were treated with PC61 at 7 and 3 days before restarting doxycycline.

Inflammatory skin disease model. To induce expression of the TGO transgene in the skin, K5/TGO/DO11 mice were maintained on 1 g kg⁻¹ doxycycline chow

(Bio-Serv). A 12-point clinical scoring scale was used to quantify skin disease. The clinical parameters of scaling, alopecia, erythema and level of activity were each given a score from 0 to 3. Scores for individual parameters were summed, and each mouse was assigned a clinical severity score out of 12. Individual mouse scores were averaged to get mean clinical scores per group.

In vitro and in vivo suppression assays. K5/TGO/DO11 mice were crossed to Foxp3-GFP transgenic mice⁹, provided by A. Rudensky. K5/TGO/DO11/Foxp3-GFP mice were started on doxycycline chow or left untreated and 6 days later, SDLN cells were isolated and purified by FACS. For *in vitro* suppression assays, cells were cultured in varying ratios with freshly isolated CFSE-labelled SDLN cells isolated from DO11.10⁺/Rag2^{-/-}/Thy1.1⁺ mice and Ova peptide-pulsed bone-marrow-derived dendritic cells. For *in vivo* suppression assays, FACS-purified T_{reg} cells were mixed 1:1 with CFSE-labelled DO11 effector cells from DO11.10⁺/Rag2^{-/-}/Thy1.1⁺ mice. Cell mixtures were injected intravenously into K5/TGO mice and recipient mice were started on doxycycline chow. Three days later, SDLN cells were isolated and CFSE dilution and intracellular cytokine expression by DO11.10⁺/Thy1.1⁺ cells was quantified by flow cytometry.

15. Saff, R. R., Spanjaard, E. S., Hohlbaum, A. M. & Marshak-Rothstein, A. Activation-induced cell death limits effector function of CD4 tumor-specific T cells. *J. Immunol.* **172**, 6598–6606 (2004).
16. Setiady, Y. Y., Coccia, J. A. & Park, P. U. *In vivo* depletion of CD4⁺FOXP3⁺ Treg cells by the PC61 anti-CD25 monoclonal antibody is mediated by FcγRIII⁺ phagocytes. *Eur. J. Immunol.* **40**, 780–786 (2010).
17. Tenorio, E. P., Fernández, J., Olguín, J. E. & Saavedra, R. Depletion with PC61 mAb before *Toxoplasma gondii* infection eliminates mainly Tregs in BALB/c mice, but activated cells in C57BL/6J mice. *FEMS Immunol. Med. Microbiol.* **62**, 362–367 (2011).

Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease

Philipp Koch^{1*}, Peter Breuer^{2*}, Michael Peitz^{1*}, Johannes Jungverdorben^{1*}, Jaideep Kesavan¹, Daniel Poppe¹, Jonas Doerr¹, Julia Ladewig¹, Jerome Mertens¹, Thomas Tüting³, Per Hoffmann⁴, Thomas Klockgether^{2,5}, Bernd O. Evert², Ullrich Wüllner² & Oliver Brüstle¹

Machado–Joseph disease (MJD; also called spinocerebellar ataxia type 3) is a dominantly inherited late-onset neurodegenerative disorder caused by expansion of polyglutamine (polyQ)-encoding CAG repeats in the *MJD1* gene (also known as *ATXN3*). Proteolytic liberation of highly aggregation-prone polyQ fragments from the protective sequence of the *MJD1* gene product ataxin 3 (*ATXN3*) has been proposed to trigger the formation of *ATXN3*-containing aggregates, the neuropathological hallmark of MJD^{1–5}. *ATXN3* fragments are detected in brain tissue of MJD patients and transgenic mice expressing mutant human *ATXN3*(Q71)⁶, and their amount increases with disease severity, supporting a relationship between *ATXN3* processing and disease progression. The formation of early aggregation intermediates is thought to have a critical role in disease initiation^{7,8}, but the precise pathogenic mechanism operating in MJD has remained elusive⁹. Here we show that L-glutamate-induced excitation of patient-specific induced pluripotent stem cell (iPSC)-derived neurons initiates Ca²⁺-dependent proteolysis of *ATXN3* followed by the formation of SDS-insoluble aggregates. This phenotype could be abolished by calpain inhibition, confirming a key role of this protease in *ATXN3* aggregation. Aggregate formation was further dependent on functional Na⁺ and K⁺ channels as well as ionotropic and voltage-gated Ca²⁺ channels, and was not observed in iPSCs, fibroblasts or glia, thereby providing an explanation for the neuron-specific phenotype of this disease. Our data illustrate that iPSCs enable the study of aberrant protein processing associated with late-onset neurodegenerative disorders in patient-specific neurons.

Despite accumulating evidence for an involvement of proteolysis in polyQ disorders, the molecular and cellular events that underlie the initial processes are still poorly understood (reviewed, for example, in ref. 10). As excitation-mediated depolarization of neurons results in transiently increased intracellular Ca²⁺ levels via influx from the extracellular space and/or the endoplasmic reticulum¹¹, we hypothesized that a temporary activation of Ca²⁺-dependent proteases might represent an early event in pathological *ATXN3* processing. To explore this hypothesis we generated neurons from iPSCs (two clones each) derived from four MJD patients (MJD1–4; for clinical details see Supplementary Fig. 1) and two related healthy donors (Ctrl1 and Ctrl2; Supplementary Fig. 1). MJD and control iPSCs exhibited expression of pluripotency genes, appropriate viral transgene silencing, potential for teratoma formation and a normal karyotype (Fig. 1a and Supplementary Figs 2–4). Expansion of the polyQ-coding CAG sequence in exon 10 of the *MJD1* gene was verified by polymerase chain reaction (PCR) analysis and genomic sequencing (Fig. 1b and Supplementary Fig. 5). From the iPSCs we generated long-term self-renewing neuroepithelial-like stem cells (lt-NES cells)¹². lt-NES cells homogeneously expressed SOX2 and, upon growth factor withdrawal,

gave rise to a dominant fraction of β III-tubulin⁺ neurons and a smaller fraction of glial fibrillary acid protein (GFAP)⁺ glia with no apparent difference between MJD and control neurons (Fig. 1c, d and Supplementary Fig. 6). lt-NES cells derived from the human embryonic stem cell line I6 (Ctrl3) and from iPSCs of a 33-year-old unrelated and healthy male donor (Ctrl4) were included as additional controls. Western blot analysis confirmed the expression from both alleles of normal *ATXN3* (~42 kDa) in control lt-NES cells and normal and expanded *ATXN3* (~60 kDa) in MJD lt-NES cells (Fig. 1e). No changes in polyQ repeat length were detected during the time course of our experiments for at least 20 passages as iPSCs, up to 47 passages as lt-NES cells and 8 weeks of neuronal differentiation (Supplementary Fig. 7). Whole-cell patch-clamp measurements of neurons differentiated for 5–6 weeks revealed the generation of multiple action potentials during current injection ($n = 6$; Fig. 1f) as well as voltage-dependent membrane currents ($n = 8$; Fig. 1g and Supplementary Fig. 6). Both excitatory and inhibitory spontaneous postsynaptic currents were observed in whole-cell voltage-clamp recordings and could be blocked by application of D(-)-2-amino-5-phosphonovaleric acid (D-APV; 50 μ M) and 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 20 μ M), or bicuculline (30 μ M), respectively ($n = 12$; Fig. 1h). No functional differences were observed between MJD and control neurons. Thus, iPSC-derived neurons are capable of generating functional neuronal networks.

To monitor excitation-induced Ca²⁺ influx into iPSC-derived neurons we loaded MJD neurons with the Ca²⁺-dependent dye Fluo-4 and stimulated the cultures with the excitatory neurotransmitter L-glutamate (100 μ M). Increased fluorescence intensity of the neurons indicated elevated excitation-mediated intracellular Ca²⁺ levels (Fig. 2a). Immunoblotting of extracts from L-glutamate-treated cultures revealed that stimulation for 30 min induced cleavage of *ATXN3* in both MJD and control neurons. A proteolytic fragment of approximately 34 kDa, corresponding to the predicted size of an amino-terminal *ATXN3* fragment was detected with an antibody to *ATXN3* (ref. 17). Stimulation with 200 μ M NMDA (*N*-methyl-D-aspartate) had a comparable effect (Fig. 2b). These results indicate that excitation-mediated Ca²⁺ influx induces cleavage of *ATXN3* in iPSC-derived neuronal cultures.

We next asked whether cleavage of *ATXN3* would also induce SDS-insoluble *ATXN3*-containing aggregates, a hallmark of MJD pathology¹⁴. To that end, lysates of neuronal cultures were fractionated to identify *ATXN3* species with different detergent solubility by immunoblotting (Fig. 2c and Supplementary Fig. 8). Lysates from control neurons prepared 30 min after stimulation contained no SDS-insoluble material, whereas lysates from MJD neurons yielded weak bands of fragmented *ATXN3* and a faint high molecular mass smear (Supplementary Fig. 9). As polyQ aggregation is associated with

¹Institute of Reconstructive Neurobiology, Life and Brain Center, University of Bonn and Hertie Foundation, 53127 Bonn, Germany. ²Department of Neurology, University of Bonn Medical Center, 53105 Bonn, Germany. ³Department of Dermatology, University of Bonn Medical Center, 53127 Bonn, Germany. ⁴Institute of Human Genetics and Department of Genomics, Life and Brain Center, University of Bonn, 53127 Bonn, Germany. ⁵DZNE, German Center for Neurodegenerative Diseases, 53175 Bonn, Germany.

*These authors contributed equally to this work.

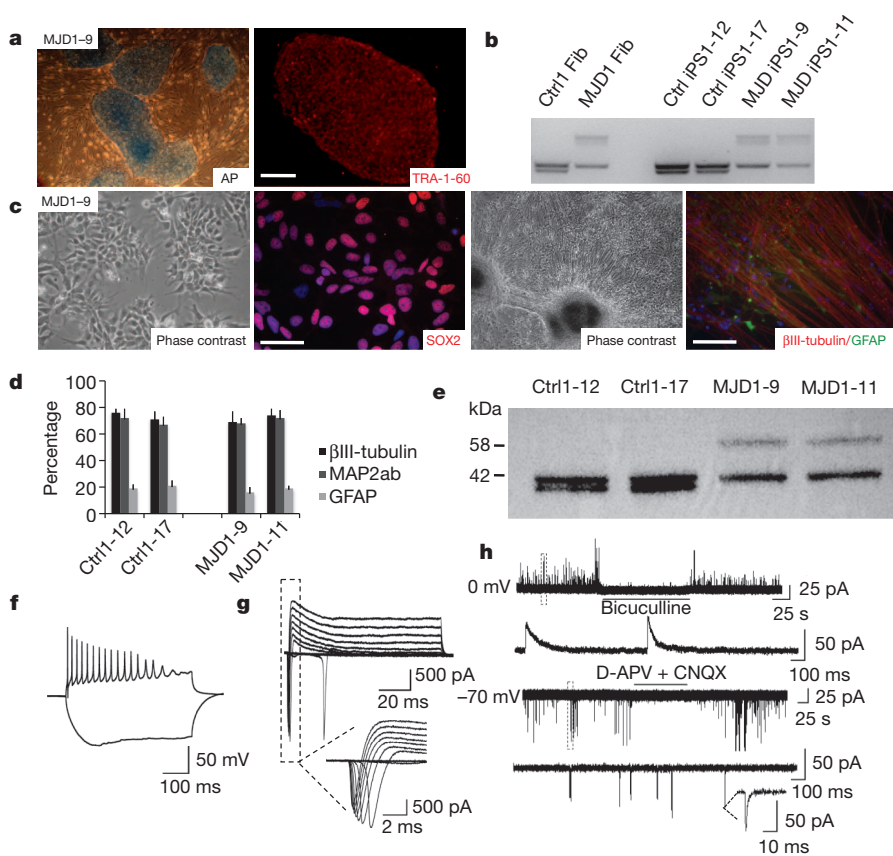


Figure 1 | Generation of patient-specific neural cultures. **a**, MJD iPSCs stain positive for alkaline phosphatase (AP) and TRA-1-60. **b**, PCR analysis to assess the length of CAG repeats. **c**, MJD iPSCs express SOX2 and differentiate into neural cultures consisting of β III-tubulin⁺ neurons and GFAP⁺ glia. **d**, Percentage of cells positive for the neuronal markers β III-tubulin and microtubulin-associated protein 2 (MAP2ab) or GFAP in 6-week-differentiated cultures. Error bars: s.d. **e**, Western blot analysis with ATXN3 antibody no. 986 (ref. 17). **f**, iPSC-derived neurons fire action potentials in response to depolarization. **g**, Depolarizing voltage steps elicited voltage-dependent fast inactivating inward and sustained outward currents. **h**, Representative traces of inhibitory (upper trace) and excitatory postsynaptic currents (bottom trace), blocked by bicuculline or D-APV and CNQX, respectively. Ctrl, control. Scale bars: **a**, 200 μ m; **c**, 50 μ m (SOX2) and 200 μ m (β III-tubulin/GFAP).

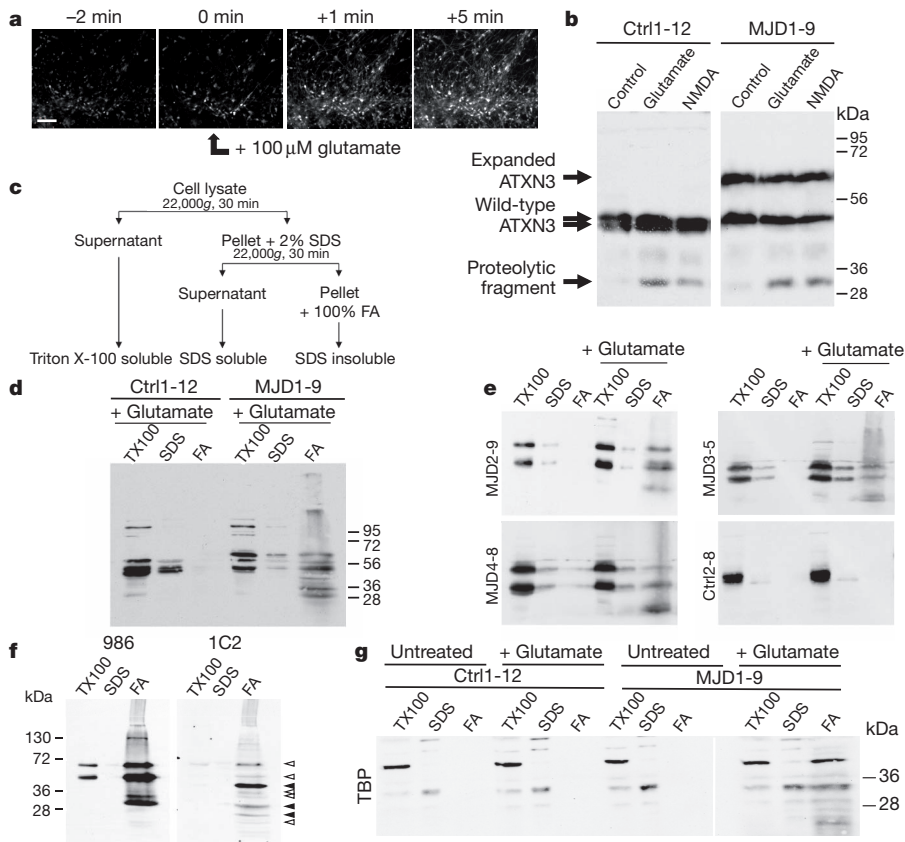


Figure 2 | Excitation induces ATXN3 cleavage and aggregation. **a**, Ca^{2+} imaging with Fluo-4 reveals a clear increase of intracellular Ca^{2+} upon exposure to L-glutamate. **b**, Exposure to L-glutamate or NMDA ($n > 3$) induces cleavage of ATXN3 (antibody no. 986). **c**, Schematic depicting the separation of cell lysates. **d**, **e**, SDS-insoluble aggregates could be robustly evoked in different MJD genotypes. Representative western blot analysis (out of $n \geq 3$) with ATXN3 antibody no. 986 (ref. 17). **f**, Re-probing of ATXN3 immunoblot (antibody no. 986; left) with the polyQ-specific 1C2 antibody (right) reveals several polyQ-containing fragments (empty and filled triangles indicate polyQ-containing fragments of which bands marked by empty triangles correspond to ATXN3 fragments). **g**, Western blot with an antibody against human TBP reveals recruitment of TBP into the SDS-insoluble fraction in MJD neurons but not control neurons ($n = 2$). FA, formic acid; TX100, Triton X-100. Scale bar: 100 μ m.

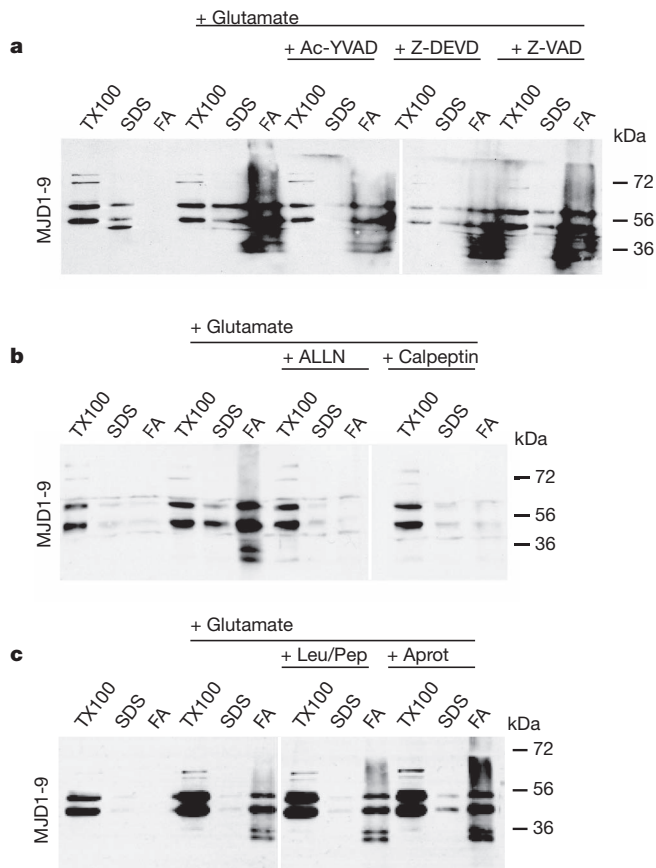


Figure 3 | Aggregation of ATXN3 is calpain dependent. **a**, ATXN3 aggregation in the presence of the caspase inhibitors Ac-YVAD-CHO, Z-DEVD-FMK or Z-VAD-FMK (all 10 μ M). **b**, ATXN3 aggregation in the presence of the calpain inhibitors ALLN (100 μ M) and calpeptin (200 μ M). **c**, ATXN3 aggregation in the presence of leupeptin/pepstatin (Leu/Pep; cysteine/aspartyl protease inhibitor) or aprotinin (Aprot; serine protease inhibitor). Shown are representative western blot analyses (out of $n > 3$) of clone MJD1-9 with ATXN3 antibody no. 986 (ref. 17). FA, formic acid; TX100, Triton X-100.

a lag phase presumably required for the generation of aggregation-competent species^{7,13,15,16}, we tested whether repetitive excitation of MJD and control neurons and a subsequent waiting period would result in more robust aggregation. Twenty-four hours after stimulation (two stimuli of 30 min) no SDS-insoluble ATXN3-positive material

was observed in control neurons. In contrast, MJD neurons now showed several prominent fragments of about 25–45 kDa and a high molecular smear, indicating the presence of fragmented and aggregated ATXN3 in these cultures (Fig. 2d). Generation of these SDS-insoluble ATXN3-containing fragments was induced in all MJD iPSC-derived neurons tested and not observed in neurons derived from healthy donors (Fig. 2e and Supplementary Fig. 10). Induction of SDS-insoluble aggregates was concentration dependent, as reduced concentrations of L-glutamate led to a reduced tendency of ATXN3 to form these aggregates (Supplementary Fig. 11). Re-probing of a typical ATXN3 immunoblot with the polyQ-specific 1C2 antibody revealed the presence of several polyQ-containing fragments within the SDS-insoluble aggregates (Fig. 2f). Mechanistically, polyQ-driven aggregation typically involves recruitment of other polyQ proteins, including the TATA binding protein (TBP), which has been detected in ATXN3-containing aggregates in MJD patients¹⁸. In agreement with these previous observations, prominent bands in the 1C2 immunoblot were identified as TBP (44 kDa) and fragments thereof (Fig. 2g and Supplementary Fig. 12). These data indicate that aggregation of ATXN3 can be robustly evoked in MJD-derived neurons and that other polyQ containing proteins are recruited into ATXN3-containing aggregates.

In MJD models, both caspase- and calpain-mediated cleavage of ATXN3 has been reported^{1,2,4,17}. However, it is still unclear which protease accounts for ATXN3 cleavage in MJD patients^{1,2,10}. As both enzymes depend on intracellular Ca^{2+} , we next tested the effect of inhibitors of both protease families on ATXN3 aggregation. No inhibition of ATXN3 aggregation was observed in cells treated with the caspase-1 inhibitor Ac-YVAD-CHO, the caspase-3 inhibitor Z-DEVD-FMK or the pan-caspase inhibitor Z-VAD-FMK (all 10 μ M; Fig. 3a and Supplementary Figs 13 and 14). In contrast, treatment with the calpain inhibitors ALLN (100 μ M) and calpeptin (200 μ M) resulted in complete disappearance of SDS-insoluble material, indicating a critical involvement of Ca^{2+} -dependent calpain proteases in the aggregation of expanded ATXN3 (Fig. 3b and Supplementary Figs 13 and 14). Inhibition of aggregate formation was also observed at lower concentrations of ALLN (10 μ M) and calpeptin (20 μ M; Supplementary Fig. 15). Inhibitors of other main protease families (200/20 μ M leupeptin/pepstatin or 2 μ M aprotinin) had no influence on aggregate formation (Fig. 3c and Supplementary Fig. 14). Thus, calpain-mediated cleavage of ATXN3 seems to be critical for aggregate formation.

Pathology in MJD patients is predominant in neurons. We thus asked whether L-glutamate-dependent aggregate formation is restricted to neurons or whether it can also be evoked in other cell types such as MJD iPSCs, MJD fibroblasts or MJD glia. Importantly, in these

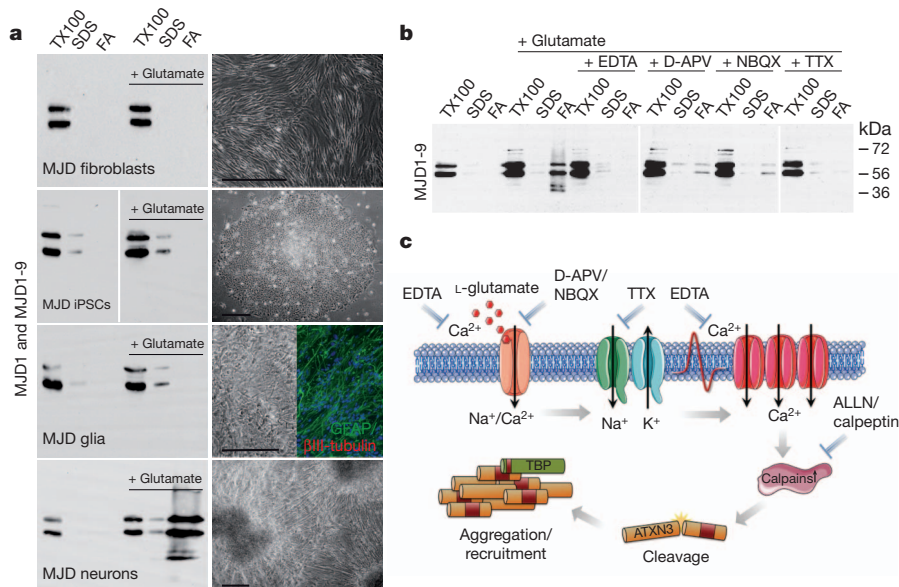


Figure 4 | ATXN3 aggregation is neuron specific. **a**, L-glutamate induces aggregation of ATXN3 in neurons but not in other cell types such as fibroblasts, iPSCs or glia. Sample 'MJD fibroblasts' was from the same experiment and processed in parallel with the other samples but blotted and exposed separately. **b**, ATXN3 aggregation in the presence of EDTA, D-APV, NBQX and TTX. Shown are representative western blot analyses (out of $n \geq 3$) of clone MJD1-9 with ATXN3 antibody no. 986 (ref. 17). **c**, Proposed model of excitation-induced aggregate formation involving activation-dependent Ca^{2+} influx via voltage-gated Ca^{2+} channels and subsequent calpain-mediated ATXN3 cleavage. Scale bars: 250 μ m. FA, formic acid; TX100, Triton X-100.

experiments only MJD neurons formed SDS-insoluble aggregates (Fig. 4a and Supplementary Fig. 14). As glial cells also express ionotropic and metabotropic glutamate receptors we became interested in whether aggregate formation in neurons involves a Ca^{2+} influx mediated by depolarization-induced activation of voltage-gated Ca^{2+} channels. Stimulation under simultaneous depletion of extracellular Ca^{2+} by EDTA (500 μM) revealed that aggregation of ATXN3 was suppressed in the presence of EDTA (Fig. 4b). Application of D-APV (50 μM) or NBQX (10 μM ; blocking ionotropic NMDA and AMPA receptors, respectively) markedly reduced the amount of aggregated ATXN3. Notably, preventing membrane-depolarization-induced activation of voltage-gated Ca^{2+} channels by blockage of Na^{+} channels with tetrodotoxin (TTX; 0.3 μM) completely inhibited the formation of aggregates (Fig. 4b and Supplementary Fig. 14). These data indicate that a neuron-specific cascade involving functional ion channels is required for protease activation and ATXN3 cleavage, and that depolarization-induced Ca^{2+} influx via voltage-gated Ca^{2+} channels activated by membrane depolarization is essential for the observed phenotype (Fig. 4c).

Interestingly, this early and clear aggregation phenotype was not accompanied by formation of ATXN3-positive inclusion bodies or macroaggregates, as analysed by epifluorescence staining 0, 1 and 24 h after stimulation, or by acute toxicity, as measured by ATP consumption and alamar blue staining 24 h, 48 h and 7 d after stimulation (Supplementary Fig. 16). This observation is not unexpected, as formation of inclusions and neuronal cell death are associated with late disease stages and are thought to be promoted by an age-related decline in protein homeostasis¹⁹. Our findings support the view that early polyQ-driven aggregation does not necessarily imply acute cytotoxicity, and that neurons can at least temporarily cope with the burden of aggregated polyQ material. Although the molecular mechanisms connecting microaggregate formation to late-stage neuronal demise remain to be elucidated, the perception that microaggregates indeed represent an early step of MJD pathology is supported by observations that they precede large inclusion bodies in several other polyQ models including SCA1, SCA7, spinobulbar muscular atrophy and other models of MJD^{20–23}. Taken together, our findings demonstrate that iPSC-derived patient-specific neurons enable the study of early aberrant protein processing in late-onset neurodegenerative disease.

METHODS SUMMARY

Cell culture. iPSCs were maintained on irradiated mouse embryonic fibroblasts (MEFs) in iPSC media containing DMEM/F12 culture medium supplemented with 20% KnockOut serum Replacement, 0.1 mM non-essential amino acids (all from Invitrogen), 1 mM L-glutamine, 0.1 mM β -mercaptoethanol and 50 ng ml⁻¹ zebrafish basic fibroblast growth factor (zbFGF). It-NES cells were established as described¹².

Excitatory stimulation and solubility analysis. MJD neurons or control neurons cultured in 3.5-cm dishes were washed three times with 2 ml BSS (balanced salt solution) containing 25 mM Tris, 120 mM NaCl, 15 mM glucose, 5.4 mM KCl, 1.8 mM CaCl_2 , 0.8 mM MgCl_2 , pH 7.4. After treatment with L-glutamate (Sigma, no. G8415) or NMDA (Sigma) in BSS for 30 min cells were washed again three times and either immediately frozen in liquid N_2 followed by lysis in RIPA buffer (50 mM Tris, 150 mM NaCl, 0.2% Triton X-100) containing 25 mM EDTA or left to recover for 30 min in differentiation media followed by a second 30 min L-glutamate treatment in BSS, and subsequently cultured in differentiation media until analysed. For inhibition studies, cells were pre-treated with the inhibitors for 30 min, followed by stimulation in the presence of the inhibitors. After treatment cells were kept in differentiation media without inhibitors until analysed. The schematic shown in Fig. 4c was produced using Servier Medical Art (<http://www.servier.com>).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 July 2010; accepted 25 October 2011.

Published online 23 November 2011.

1. Berke, S. J., Schmied, F. A., Brunt, E. R., Ellerby, L. M. & Paulson, H. L. Caspase-mediated proteolysis of the polyglutamine disease protein ataxin-3. *J. Neurochem.* **89**, 908–918 (2004).

2. Jung, J., Xu, K., Lessing, D. & Bonini, N. M. Preventing ataxin-3 protein cleavage mitigates degeneration in a *Drosophila* model of SCA3. *Hum. Mol. Genet.* **18**, 4843–4852 (2009).
3. Tarlac, V. & Storey, E. Role of proteolysis in polyglutamine disorders. *J. Neurosci. Res.* **74**, 406–416 (2003).
4. Wellington, C. L. et al. Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J. Biol. Chem.* **273**, 9158–9167 (1998).
5. Rubinsztein, D. C., Wyttenbach, A. & Rankin, J. Intracellular inclusions, pathological markers in diseases caused by expanded polyglutamine tracts? *J. Med. Genet.* **36**, 265–270 (1999).
6. Goti, D. et al. A mutant ataxin-3 putative-cleavage fragment in brains of Machado-Joseph disease patients and transgenic mice is cytotoxic above a critical concentration. *J. Neurosci.* **24**, 10266–10279 (2004).
7. Chen, S., Ferrone, F. A. & Wetzel, R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl Acad. Sci. USA* **99**, 11884–11889 (2002).
8. Schaffar, G. et al. Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. *Mol. Cell* **15**, 95–105 (2004).
9. Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Rev. Genet.* **6**, 743–755 (2005).
10. Williams, A. J. & Paulson, H. L. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* **31**, 521–528 (2008).
11. Berridge, M. J. Neuronal calcium signaling. *Neuron* **21**, 13–26 (1998).
12. Koch, P., Opitz, T., Steinbeck, J. A., Ladewig, J. & Brüstle, O. A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for *in vitro* instruction and synaptic integration. *Proc. Natl Acad. Sci. USA* **106**, 3225–3230 (2009).
13. Thakur, A. K. & Wetzel, R. Mutational analysis of the structural organization of polyglutamine aggregates. *Proc. Natl Acad. Sci. USA* **99**, 17014–17019 (2002).
14. Ikeda, H. et al. Expanded polyglutamine in the Machado-Joseph disease protein induces cell death *in vitro* and *in vivo*. *Nature Genet.* **13**, 196–202 (1996).
15. Scherzinger, E. et al. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates *in vitro* and *in vivo*. *Cell* **90**, 549–558 (1997).
16. Haacke, A. et al. Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. *Hum. Mol. Genet.* **15**, 555–568 (2006).
17. Haacke, A., Hartl, F. U. & Breuer, P. Calpain inhibition is sufficient to suppress aggregation of polyglutamine-expanded ataxin-3. *J. Biol. Chem.* **282**, 18851–18856 (2007).
18. Uchiyama, T. et al. Non-expanded polyglutamine proteins in intranuclear inclusions of hereditary ataxias: triple-labeling immunofluorescence study. *Acta Neuropathol.* **102**, 149–152 (2001).
19. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
20. Yoo, S. Y. et al. SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity. *Neuron* **37**, 383–401 (2003).
21. Watase, K. et al. A long CAG repeat in the mouse *Scal* locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* **34**, 905–919 (2002).
22. Li, M., Chevalier-Larsen, E. S., Merry, D. E. & Diamond, M. I. Soluble androgen receptor oligomers underlie pathology in a mouse model of spinobulbar muscular atrophy. *J. Biol. Chem.* **282**, 3157–3164 (2007).
23. Williams, A. J., Knutson, T. M., Colomer-Gould, V. F. & Paulson, H. L. *In vivo* suppression of polyglutamine neurotoxicity by C-terminus of Hsp70-interacting protein (CHIP) supports an aggregation model of pathogenesis. *Neurobiol. Dis.* **33**, 342–353 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Auel, D. Kühne, M. Segschneider and V. Dobberahn for technical support, A. Leinhaas for conducting the teratoma assays and H. L. Paulson for providing the polyclonal ATXN3 antibody used for immunocytochemistry. The human embryonic stem cell lines H9 and I3 (used as control for quantitative PCR) and I6 (originally used to derive the It-NES cells) were provided by J. Itskovitz-Eldor (Technion, Israel Institute of Technology, Haifa, Israel). The work was supported by the German Federal Ministry for Education and Research (BMBF; grants 01GN0813, 01GS0860), the European Union (LSHG-CT-2006-018739, ESTOOLS; HEALTH-F5-2010-266753, SCR&Tox), the Deutsche Forschungsgemeinschaft (WU184/6-1, EV143/1-1), BONFOR and the Hertie Foundation.

Author Contributions P.K. and P.B.: conception and study design, directed differentiation of iPSCs, cellular/molecular assays for disease modelling, data assembly, analysis and interpretation, writing of manuscript; M.P. and J.J.: iPSC clone derivation, maintenance and validation, directed differentiation of iPSCs; J.K., D.P., J.D., J.L., J.M., P.H.: data collection, analysis and interpretation; B.O.E., U.W., T.K.: conception, data analysis and interpretation; T.T.: provision of material; O.B.: conception, data analysis and interpretation, writing of manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.B. (brustle@uni-bonn.de).

METHODS

Retrovirus production and human iPSC induction. Retroviral plasmids for OCT4, SOX2, KLF4 and c-MYC²⁴ (plasmids 17217, 17218, 17219 and 17220) were obtained from Addgene. PhoenixGP cells (ATCC no. 3514) cultured in 10-cm plates were co-transfected with retroviral and VSV-G vector (plasmid 12259, Addgene) using FUGENE HD (Roche). Viral supernatants were harvested on day two and three after transfection, filtered through a 0.45- μ m cellulose acetate filter and centrifuged at 5×10^4 average g force for 90 min. Dermal fibroblasts were plated at a density of 5×10^4 cells per well of a 12-well plate and infected overnight with retrovirus in the presence of 6 μ g ml⁻¹ polybrene and 4 ng ml⁻¹ FGF2. The infection procedure was repeated the next day. For the reprogramming of MJD2, MJD3, MJD4 and Ctrl2 alternative retroviruses coding for OCT4, SOX2, KLF4 and c-MYC (Addgene plasmids 20072, 20073, 20074 and 20075) were produced in Plata cells (Cell Biolabs). Four days after transduction, cells were split into plates pre-seeded with mouse embryonic fibroblasts (MEFs). Medium was switched to human iPSC medium containing DMEM/F12 culture medium supplemented with 20% KnockOut Serum Replacement, 0.1 mM non-essential amino acids (all from Invitrogen), 1 mM L-glutamine, 0.1 mM β -mercaptoethanol and 50 ng ml⁻¹ zebrafish basic fibroblast growth factor (zbFGF) 6 days after transduction and changed every other day. Four weeks after transduction, colonies with human ES cell morphology were manually picked, carefully triturated and expanded to establish human iPSC lines. iPSCs were maintained on irradiated MEFs in iPSC medium and split at a 1:3 to 1:4 ratio every 4–5 days.

Teratoma formation. For teratoma formation, 1×10^6 cells were resuspended as clumps in human ES cell medium and injected into the testis of immune-compromised SCID beige mice. After 5–8 weeks, tumours were isolated, fixed overnight in 4% paraformaldehyde (PFA) and subjected to histological examination using haematoxylin and eosin staining.

SNP analysis. Genomic DNA was prepared using the DNeasy Blood & Tissue Kit (Qiagen). Whole-genome single nucleotide polymorphism (SNP) genotyping was performed at the Institute of Human Genetics at the University of Bonn. Genomic DNA at a concentration of 50 ng μ l⁻¹ was used for whole-genome amplification. Afterwards, the amplified DNA was fragmented and hybridized to sequence-specific oligomers bound to beads on an Illumina Human610-Quad chip or a HumanCytoSNP-12 chip. Data were analysed using Illumina BeadStudio.

Analysis of transgene silencing. Total RNA was extracted using TRIzol reagent (Invitrogen). To generate cDNA, RNA was incubated with DNase I (Promega) and reverse transcribed with BioRad iScript Kit (BioRad) according to the manufacturer's manual. Quantitative RT-PCR was performed using the iCycler (BioRad) platform following the manufacturer's instruction. All results were normalized to GAPDH. Primers used for the analysis of transgene silencing and expression of pluripotency-associated genes were: OCT4 endo, forward, GACAGGGGGAGGGGAGGAGCTAG; reverse, GTTCCCTCCAACCAGTTGCCCCAAC; OCT4 total, forward, GTGGAGGAAGCTGACAACA; reverse, ATTCTCCAGGTTGCCTCTCA; SOX2 endo, forward, GTATCAGGAGTTGTC AAGGCAGAG; reverse, TCCTAGTCTTAAAGAGGCAGCAAC; SOX2 total, forward, CCCGAGTGGAACTTTTGTCG; reverse, GCAGCGTGACTTA TCCTTCTT; c-MYC endo, forward, TTCGGGTAGTGGAAAACAG; reverse, CCTCTCGTCGCAGTAGAAA; c-MYC total, forward, AAGACTCCAGCG CCTTCTCT; reverse, TCTTGTTCCTCCTCAGAGTCG; KLF4 endo, forward, GACCAGGCACTACCGTAAACA; reverse, CTGGCAGTGTGGGTCATATC; KLF4 total, forward, CCCAATTACCATCCTTCTCT; reverse, ACGATCGTCTT CCCCTCTT; GAPDH, forward, TGACAACTTTGGTATCGTGGA; reverse, CAGTAGAGGCAGGGATGAT. 'Total' refers to primers for amplification of transgenes and endogenous genes, 'endo' to primers for amplification of endogenous genes only.

Genomic repeat length analysis. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). PCR amplification of exon 10 of the *MJD1* gene was performed using Phusion Polymerase (Finnzymes; primers: MJDFw, AGCACTT CCATATTTTAAAGTAATCTG; MJDr, TGCTCCTTAATCCAGGGAAA). PCR included a pre-incubation at 98 °C for 30 s followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s and extension at 72 °C for 60 s with a final extension step at 72 °C for 10 min. Fragments were extracted from the gel and subsequently sequenced.

Neural differentiation of iPSCs. iPSC colonies were detached with collagenase (Invitrogen) and plated on non-adhesive plastic dishes in iPSC media without FGF2 and β -mercaptoethanol to form embryoid bodies (EBs). Medium was changed every second day, and at day 5–7 EBs were plated on poly-L-ornithine/laminin (both Sigma) coated tissue culture plates. Neural rosette structures started to emerge about 1 week after plating and were carefully picked at day 8–12. Picked neural rosettes were cultured in suspension for 2 days in DMEM/F12, 2 mM L-glutamine, 1.6 g l⁻¹ glucose, 0.1 mg ml⁻¹ penicillin/streptomycin, N2 supplement (1:100; Invitrogen), then dissociated with trypsin, plated onto poly-L-ornithine/laminin-coated plates

and propagated in N2 medium supplemented with B27 (1 μ l ml⁻¹, Invitrogen), 10 ng ml⁻¹ FGF2 and 10 ng ml⁻¹ EGF (both from R&D systems). Cells were passaged at a ratio of 1:2–1:3 every 2–3 days using trypsin. Neuronal differentiation was induced as described¹² by removing the growth factors FGF2 and EGF from the media and culturing the cells in Neurobasal medium supplemented with B27 (1:50, Invitrogen) and DMEM/F12 supplemented with N2 (1:100) mixed at a 1:1 ratio. 300 ng ml⁻¹ cAMP was added to the differentiation media. Stimulation experiments were conducted in neuronal cultures 6–8 weeks after growth factor withdrawal. Differentiation into glial cells was performed by culturing MJD It-NES cells in N2 medium supplemented with 5% fetal calf serum (FCS) and 10 ng ml⁻¹ CNTF. Three weeks later cells were trypsinized, filtered through a cell mesh and cultured in N2 medium supplemented with 5% FCS and 10 ng ml⁻¹ EGF for 3–5 passages. Glial cultures used for the experiments were stained for GFAP and β III-tubulin in parallel and showed only very occasional β III-tubulin positive neurons (<1:1,000).

Immunocytochemical analysis. Cells were fixed with 4% PFA in PBS for 10 min at room temperature (20 °C) and blocked in 10% FCS (Invitrogen) in PBS. For detection of the pluripotency-associated markers TRA-1-60, TRA-1-81, SSEA-3 and SSEA-4 samples were incubated with primary antibodies at room temperature for 2 h, washed three times, incubated with secondary antibody for 45 min, counterstained with DAPI and mounted with Mowiol 4-88 mounting solution (Carl Roth). For all other antibodies cells were blocked in 0.1% Triton X-100 (Sigma) and 10% FCS in PBS, incubated with the primary antibodies for 16 h at 4 °C, washed three times, incubated with secondary antibody, counterstained with DAPI and mounted with Mowiol 4-88.

Primary antibodies and concentrations were as follows: TRA-1-60 (1:500, Invitrogen), TRA-1-81 (1:500, Invitrogen), SSEA3 and SSEA4 (both 1:80, gifts from P. Andrews; ESTOOLS consortium), SOX2 (1:500, R&D Systems), nestin (1:600, R&D Systems), DACH1 (1:100, Proteintech), PLZF (1:50, Calbiochem), ZO-1 (1:100, Zymed), β III-tubulin (1:2,500, Covance), MAP2ab (1:250, Chemicon), GFAP (1:1,000, DAKOCytomation), GABA (1:800, Sigma) and ATXN3 (1:1,000, gift from H. L. Paulson). Secondary antibodies were Alexa488 anti-ms, Alexa555 anti-ms, Alexa488 anti-rb, Alexa555 anti-rb (all 1:1,000, Invitrogen), Cy3 anti-rat (1:300, Jackson/Dianova). Images were acquired with an AxioCam MRn and processed using Axiovision Software (Zeiss).

Western blotting. MJD neurons or control neurons cultured in 3.5-cm dishes were washed three times with 2 ml BSS (balanced salt solution) containing 25 mM Tris, 120 mM NaCl, 15 mM glucose, 5.4 mM KCl, 1.8 mM CaCl₂, 0.8 mM MgCl₂, pH 7.4. After treatment with L-glutamate (Sigma, no. G8415) or NMDA (Sigma) in BSS for 30 min cells were washed again three times and either immediately frozen in liquid N₂ followed by lysis in RIPA buffer (50 mM Tris, 150 mM NaCl, 0.2% Triton X-100) containing 25 mM EDTA or left to recover for 30 min in differentiation media followed by a second 30 min L-glutamate treatment in BSS, and subsequently cultured in differentiation media until analysed. For inhibition studies, cells were pre-treated with the inhibitors for 30 min, followed by stimulation in the presence of the inhibitors. After treatment cells were kept in differentiation media without inhibitors until analysed. For analysis of fragmentation and aggregation of ATXN3 by SDS-PAGE and western blotting, extracts were analysed either immediately after lysis or after fractionation. For fractionation, lysates containing 1–2 μ g μ l⁻¹ total protein dissolved in 50 mM Tris, 150 mM NaCl, 0.2% Triton X-100, 25 mM EDTA (RIPA buffer) were centrifuged at 22,000g for 30 min at 4 °C. The pellet fractions were separated from supernatants (Triton X-100-soluble fraction) and homogenized in 150 μ l RIPA buffer containing 2% SDS followed by a second centrifugation step at room temperature. The supernatants (SDS-soluble fraction) were removed, and the remaining pellets were incubated for 16 h in 100% formic acid at 37 °C²⁵. After evaporation in a Speed Vac concentrator (Thermo Scientific), the pellet was dissolved in 50 μ l Laemmli-buffer (SDS-insoluble fraction) followed by pH adjustment with 2 M Tris-base for SDS-PAGE analysis. Gels were loaded with 40 μ g of the Triton X-100 fraction, 37.5 μ l of the SDS-soluble fraction and the complete SDS-insoluble fraction. Western blot analyses were performed following standard procedures¹⁷. Before probing membranes with different antibodies, bound antibodies were removed by incubation in stripping buffer (63 mM Tris/HCl (pH 6.7), 2% SDS, 100 mM β -mercaptoethanol) for 1 h at 50 °C followed by extensive washes in TBS. Extracts for assessing glutamate stimulation of iPSCs were prepared from feeder-free cultures. The raise and purification of the polyclonal antiserum against rat ATXN3 (no. 986) is described elsewhere¹⁷. The monoclonal anti-polyQ antibody (clone 1C2) and the monoclonal anti-TBP antibody to 18 N-terminal amino acid residues of TBP (MAB 3658) were purchased from Chemicon, the anti-actin antibody (AC40) was obtained from Sigma, and anti-Ach3 antibody from Upstate. Unless specified otherwise in the figure legends, compiled blots in the individual figures represent samples from the same experiment, which were processed, blotted and exposed in parallel.

Electrophysiological recordings. Whole-cell current-clamp and voltage-clamp recording was carried out with Axopatch-200B amplifier (Molecular Devices) that was interfaced by an A/D-converter (Digidata 1440, Molecular Devices) to a PC

running PClamp software (Version 10, Molecular Devices). Pipette electrodes (GB150F-8P, Science Products) were fabricated using a vertical puller (Narishige PC-10) and fire polished (final tip resistances 2–4 M Ω). All recordings were performed at room temperature in a bath solution containing (in mM): 140 NaCl, 5 KCl, 2 CaCl₂, 1 MgCl₂, 10 HEPES and 25 D-glucose (pH 7.2, osmolality 310–320 mOsm). For most recordings of membrane potential or current, the patch pipette contained the following (in mM): 120 potassium gluconate (C₆H₁₁O₇K) or potassium fluoride, 20 KCl, 10 NaCl, 10 EGTA, 1 CaCl₂, 4 Mg ATP, and 0.4 Na GTP and 10 HEPES (pH 7.2, osmolality 280–290 mOsm). For some voltage-clamp recording, another pipette solution was used (in mM): 110 cesium fluoride, 11 EGTA, 5 ATP-Na, 0.5 GTP-Tris, 2 MgCl₂, 20 TEA-Cl and 10 HEPES-Na (pH 7.2, osmolality 280–290 mOsm). To isolate excitatory and inhibitory postsynaptic currents (sEPSCs and sIPSCs), another pipette filling solution was used (in mM): 110 cesium methanesulphonate (CH₃O₃SCs), 10 TEA-Cl, 5 QX-314, 10 EGTA, 1 CaCl₂, 10 phosphocreatine Na₂, 4 Mg-ATP, 0.4

Na-GTP and 0.25% biocytin (pH 7.2, 280–290 mOsm). For this solution holding potential was corrected for 8 mV junction potential. sEPSCs and sIPSCs were recorded at a V_{hold} of –70 mV and 0 mV, respectively.

Survival assays. MJD neurons or Ctrl neurons were plated into 96-well dishes and differentiated for 6 weeks in differentiation media. Cells were stimulated twice with glutamate as described earlier and analysed 24 h, 48 h and 7 days later. Analysis of alamar blue staining and quantification of the ATP content was performed using the commercial kits alamar blue (Invitrogen) and CellTiter-Glo (Promega) according to the manufacturer's instructions.

24. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
25. Hazeki, N., Tukamoto, T., Goto, J. & Kanazawa, I. Formic acid dissolves aggregates of an N-terminal huntingtin fragment containing an expanded polyglutamine tract: applying to quantification of protein components of the aggregates. *Biochem. Biophys. Res. Commun.* **277**, 386–393 (2000).

Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease

Sonja Kriks^{1,2*}, Jae-Won Shim^{1,2*}, Jinghua Piao^{1,3}, Yosif M. Ganat^{1,2}, Dustin R. Wakeman⁴, Zhong Xie⁵, Luis Carrillo-Reid⁵, Gordon Auyeung^{1,3}, Chris Antonacci^{1,3}, Amanda Buch^{1,3}, Lichuan Yang⁶, M. Flint Beal⁶, D. James Surmeier⁵, Jeffrey H. Kordower⁴, Viviane Tabar^{1,3} & Lorenz Studer^{1,2,3}

Human pluripotent stem cells (PSCs) are a promising source of cells for applications in regenerative medicine. Directed differentiation of PSCs into specialized cells such as spinal motoneurons¹ or midbrain dopamine (DA) neurons² has been achieved. However, the effective use of PSCs for cell therapy has lagged behind. Whereas mouse PSC-derived DA neurons have shown efficacy in models of Parkinson's disease^{3,4}, DA neurons from human PSCs generally show poor *in vivo* performance⁵. There are also considerable safety concerns for PSCs related to their potential for teratoma formation or neural overgrowth^{6,7}. Here we present a novel floor-plate-based strategy for the derivation of human DA neurons that efficiently engraft *in vivo*, suggesting that past failures were due to incomplete specification rather than a specific vulnerability of the cells. Midbrain floor-plate precursors are derived from PSCs 11 days after exposure to small molecule activators of sonic hedgehog (SHH) and canonical WNT signalling. Engraftable midbrain DA neurons are obtained by day 25 and can be maintained *in vitro* for several months. Extensive molecular profiling, biochemical and electrophysiological data define developmental progression and confirm identity of PSC-derived midbrain DA neurons. *In vivo* survival and function is demonstrated in Parkinson's disease models using three host species. Long-term engraftment in 6-hydroxy-dopamine-lesioned mice and rats demonstrates robust survival of midbrain DA neurons derived from human embryonic stem (ES) cells, complete restoration of amphetamine-induced rotation behaviour and improvements in tests of forelimb use and akinesia. Finally, scalability is demonstrated by transplantation into parkinsonian monkeys. Excellent DA neuron survival, function and lack of neural overgrowth in the three animal models indicate promise for the development of cell-based therapies in Parkinson's disease.

Recent mouse genetic studies have demonstrated an important role for the transcription factor *FOXA2* during midbrain DA neuron development^{8,9}. A unique feature of the developing midbrain is the coexpression of the floor-plate (FP) marker *FOXA2* and the roof plate marker *LMX1A*. Normally, FP and roof plate cells are located at distinct positions in the central nervous system (ventral versus dorsal) and show diametrically opposed patterning requirements^{10,11}. We recently reported the derivation of FP precursors from ES cells¹² using a modified dual-SMAD inhibition protocol¹³. Canonical Wnt signalling is important for both roof plate function¹⁴ and midbrain DA neuron development¹⁵. We therefore proposed that WNT activation may induce *LMX1A* expression and neurogenic conversion of PSC-derived midbrain FP towards DA neuron fate. Here we report that exposure to CHIR99021 (CHIR), a potent GSK3B inhibitor known to

strongly activate WNT signalling¹⁶, induces *LMX1A* in *FOXA2*+ FP precursors (Fig. 1a). CHIR was much more potent than recombinant WNT3A or WNT1 at inducing *LMX1A* expression (data not shown). The efficiency of *LMX1A* induction was dependent on the timing of CHIR exposure with a maximum effect from day 3 to day 11 (Supplementary Fig. 1). CHIR was the most critical factor for inducing coexpression of *FOXA2/LMX1A*, while other factors such as FGF8 had only marginal effects (Supplementary Fig. 2). Induction of *FOXA2/LMX1A* coexpression required strong activation of SHH signalling using purmorphamine, a small molecule agonist, alone or in combination with recombinant SHH (Supplementary Fig. 3). Treatment with SHH agonists (purmorphamine + SHH) and FGF8 (S/F8) in the absence of CHIR showed significantly lower expression of *FOXA2* by day 11 and complete lack of *LMX1A* expression (Fig. 1a, b). Dual SMAD inhibition (exposure to LDN193189 + SB431542, 'LSB') did not yield *FOXA2*-expressing cells, but a subset of *LMX1A*+ cells (Fig. 1a, b). The anterior marker *OTX2* was robustly induced in LSB- and LSB/S/F8/CHIR-treated cultures, but not under LSB/S/F8 conditions (Fig. 1a, c). Systematic comparisons of the three culture conditions (Fig. 1d) were performed using global temporal gene expression profiling. Hierarchical clustering of differentially expressed genes segregated the three treatment conditions by day 11 of differentiation (Supplementary Fig. 4a). *FOXA1*, *FOXA2* and several other SHH downstream targets including *PTCH1* were among the most differentially regulated transcripts in LSB/S/F8/CHIR versus LSB treatment sets (Fig. 1e). Expression of *LMX1A*, *NGN2* (also known as *NEUROG2*) and *DDC* indicated establishment of midbrain DA neuron precursor fate already by day 11 (Fig. 1e, f). In contrast, LSB cultures were enriched for dorsal forebrain precursor markers such as *HES5*, *PAX6*, *LHX2* and *EMX2*. Direct comparison of LSB/S/F8/CHIR versus LSB/S/F8 treatment (Fig. 1f) confirmed selective enrichment for midbrain DA precursor markers in LSB/S/F8/CHIR group and suggested hypothalamic precursor identity in LSB/S/F8-treated cultures based on the differential expression of *RAX*, *SIX3* and *SIX6* (ref. 17) (see also *POMC* and *OTP* expression in Fig. 2d below). The full list of differentially expressed transcripts (Supplementary Tables 1 and 2) and gene ontology analysis (Supplementary Fig. 4b) (DAVID; <http://david.abcc.ncifcrf.gov>, ref. 18) confirmed enrichment for canonical WNT signalling upon CHIR treatment (raw data available at the Gene Expression Omnibus). Comparative temporal gene expression analysis for markers of midbrain DA precursors (Fig. 1g) versus anterior and ventral non-DA fates (Fig. 1h) partitioned the three induction conditions into (1) LSB: dorsal forebrain; (2) LSB/S/F8: ventral/hypothalamic; (3) LSB/S/F8/CHIR: midbrain DA identity.

¹Center for Stem Cell Biology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ²Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ³Department of Neurosurgery, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ⁴Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois 60612, USA. ⁵Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. ⁶Department of Neurology and Neuroscience, Weill Medical College of Cornell University, New York Presbyterian Hospital, 525 East 68th Street, New York, New York 10021, USA.

*These authors contributed equally to this work.

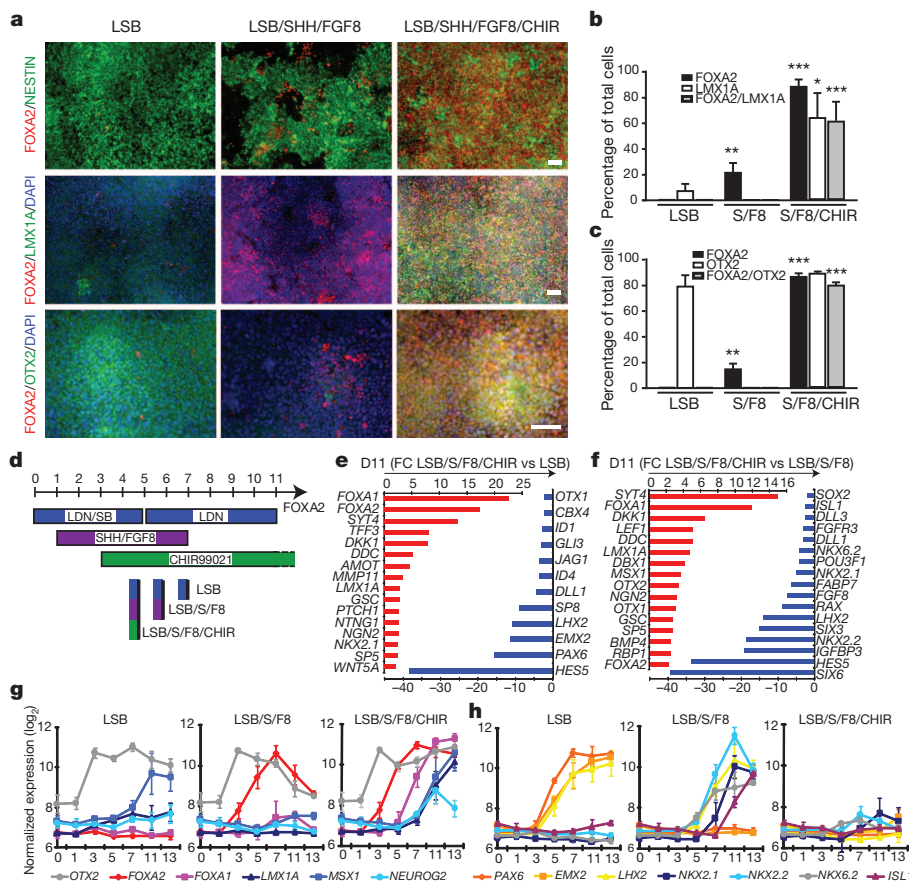


Figure 1 | Induction and neurogenic conversion of ES-cell-derived midbrain FP precursors is dependent on CHIR addition.

a, Immunocytochemistry at day 11 for FOXA2 (red), NESTIN (green, upper panels), LMX1A (green, middle panels) and OTX2 (green, lower panels). **b, c**, Quantification of the data presented in **a**; mean \pm s.e.m., $n = 3$ (independent experiments); *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$ (compared to LSB, Dunnett test). **d**, Diagram of culture conditions. **e, f**, Selected lists of differentially expressed transcripts at day 11 comparing LSB/S/F8/CHIR versus LSB (**e**) or versus LSB/S/F8 (**f**). **g, h**, Temporal gene expression analysis for markers of midbrain DA precursor identity (**g**), forebrain and ventral non-DA precursor identity (**h**). Scale bars, 50 μ m.

By day 25, all three conditions yielded TUJ1+ neurons (Fig. 2a) and cells expressing TH, the rate-limiting enzyme in the synthesis of DA. However, only the LSB/S/F8/CHIR group yielded TH+ cells coexpressing LMX1A and FOXA2 as well as the nuclear receptor NURR1 (also

known as NR4A2) (Fig. 2a, b). Comparing gene expression in day 13 versus day 25 cultures confirmed robust induction of postmitotic DA neuron markers (Fig. 2c). Characterizing DA neuron identity at day 25 in comparison to LSB and LSB/S/F8 showed enrichment for known

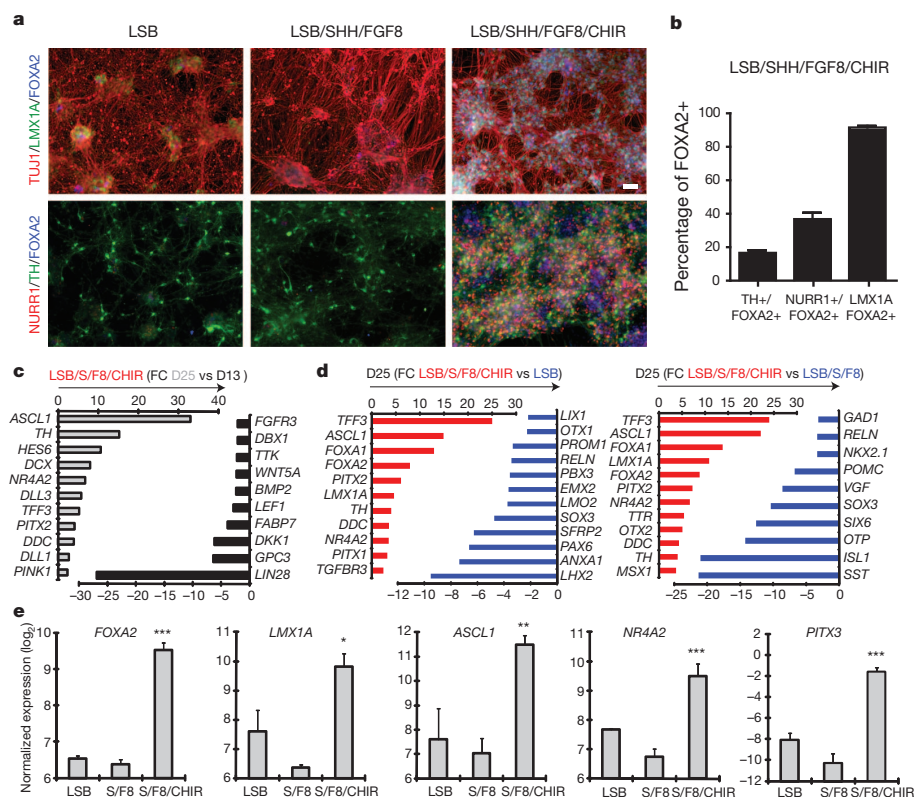


Figure 2 | Immunocytochemical and molecular analysis of midbrain DA neuron fate in LSB/S/F8/CHIR-treated versus LSB/S/F8 (hypothalamic) and forebrain LSB (dorsal forebrain) fates. **a**, Immunocytochemistry at day 25 for coexpression of FOXA2 (blue) with Tuj1 (red)/LMX1A (green) (upper panels) and NURR1 (red)/TH (green) (lower panels). **b**, Quantitative coexpression analysis for LSB/S/F8/CHIR; mean \pm s.e.m., $n = 3$ (independent experiments). **c, d**, Global gene expression analysis at day 25 (quadruplicates each). Selected lists of differentially expressed transcripts comparing day 13 versus day 25 in LSB/S/F8/CHIR (**c**) LSB/S/F8/CHIR versus LSB (**d**, left panel) and LSB/S/F8 (**d**, right panel). **e**, Gene expression analysis for key midbrain DA neuron markers. Significance compared to LSB (Dunnett test): *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. Scale bars, 50 μ m.

a, Immunocytochemistry at day 25 for coexpression of FOXA2 (blue) with Tuj1 (red)/LMX1A (green) (upper panels) and NURR1 (red)/TH (green) (lower panels). **b**, Quantitative coexpression analysis for LSB/S/F8/CHIR; mean \pm s.e.m., $n = 3$ (independent experiments). **c, d**, Global gene expression analysis at day 25 (quadruplicates each). Selected lists of differentially expressed transcripts comparing day 13 versus day 25 in LSB/S/F8/CHIR (**c**) LSB/S/F8/CHIR versus LSB (**d**, left panel) and LSB/S/F8 (**d**, right panel). **e**, Gene expression analysis for key midbrain DA neuron markers. Significance compared to LSB (Dunnett test): *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. Scale bars, 50 μ m.

midbrain DA neuron transcripts and identified multiple novel candidate markers (Fig. 2d, Supplementary Tables 3–5 and Supplementary Fig. 4b). For example, the transcript most highly enriched in LSB/S/F8/CHIR (midbrain DA group) was *TFF3*, a gene not previously associated with midbrain DA neuron development, but highly expressed in the human substantia nigra (Supplementary Fig. 4c; Allen Brain Atlas <http://human.brain-map.org>). Similar data were obtained for *EBF1*, *EBF3* (Supplementary Fig. 4c) as well as *TTR*, a known transcriptional target of *FOXA2* in the liver¹⁹. We observed enrichment of several *PITX* genes, and *PITX3*, a classic marker of midbrain DA neurons, was also expressed robustly at day 25 of differentiation (Fig. 2e). Finally, both midbrain FP and DA neuron induction was readily reproduced in independent ES cell and human induced PSC lines (Supplementary Fig. 5). Our data demonstrate that the LSB/S/F8/CHIR protocol yields cells expressing a marker profile matching midbrain DA neuron fate.

We next proceeded to determine the *in vitro* and *in vivo* properties of FP-derived DA neurons in comparison to DA neurons obtained via a neural rosette intermediate³ (Supplementary Fig. 6). Patterning of neural rosettes represents the currently most widely used strategy for deriving DA neurons from PSCs^{2,6,20}. Both FP- and rosette-based protocols were efficient at generating TH⁺ neurons capable of long-term *in vitro* survival (day 50 of differentiation; Fig. 3a). However, the percentage of TH⁺ cells was significantly higher in FP-derived cultures (Fig. 3b). Whereas TH⁺ cells in both protocols showed co-expression of *NURR1*, only FP-derived DA neurons coexpressed *FOXA2* and *LMX1A* (Fig. 3a,b). Few GABA- (γ -aminobutyric acid) and serotonin (5-HT)-positive neurons were observed (Fig. 3c). DA, and its metabolites DOPAC (3,4-dihydroxy-phenylacetic acid) and HVA (homovanillic acid), were present in cultures generated with either protocol, but DA levels were approximately eight times higher in FP cultures (Fig. 3d, e). Midbrain DA neurons showed extensive fibre outgrowth and robust expression of mature neuronal markers including synapsin, dopamine transporter (DAT) and G-protein-coupled, inwardly rectifying potassium channel (*KCNJ6* (*Kir3.2*), also called *GIRK2*), expressed in substantia nigra pars compacta (SNPC) DA neurons (Fig. 3f and Supplementary Fig. 7).

SNPC DA neurons *in vivo* show an electrophysiological phenotype that differentiates them from most other neurons in the brain. In particular, they spike spontaneously at a slow (1–3 Hz) rate. Moreover, this slow spiking is accompanied by a slow, sub-threshold oscillatory potential^{21,22}. After 2–3 weeks *in vitro*, these same physiological features are shown by SNPC DA neurons grown from early postnatal mice (data not shown). The DA neurons differentiated from ES cells consistently (4/4) had this distinctive physiological phenotype (Fig. 3g–i). Future studies will be required to determine whether all features of SNPC DA neurons are recapitulated by the ES cell DA neurons *in vitro* or whether full differentiation will require maturation *in vivo*. However, our data indicate that FP-derived DA neurons do show the cardinal physiological features of mature SNPC DA neurons.

A major challenge in the field has been the ability to generate PSC-derived midbrain DA neurons that functionally engraft *in vivo* without the risk of neural overgrowth⁷ or inappropriate differentiation into non-midbrain neurons²³. On the basis of fetal tissue transplantation studies²⁴ we proposed that the time of cell cycle exit, marked by expression of *NURR1* (ref. 25), may be a suitable stage for grafting (~day 25 of differentiation, Fig. 2). Initial studies using day 25 cells in non-lesioned adult mice showed robust survival of PSC-derived *FOXA2*⁺/TH⁺ neurons at 6 weeks after transplantation (Supplementary Fig. 8). We next addressed whether *FOXA2*⁺/TH⁺ cells survive long-term in parkinsonian hosts without resulting in neural overgrowth. To this end, we made 6-hydroxy-dopamine (6-OHDA) lesions³ in *NOD-SCID IL2Rgc* null mice, a strain that efficiently supports xenograft survival with particular sensitivity for exposing rare tumorigenic cells²⁶. Both FP- and rosette-derived cultures were grafted (1.5×10^5 per animal) without prior purification to reveal potential

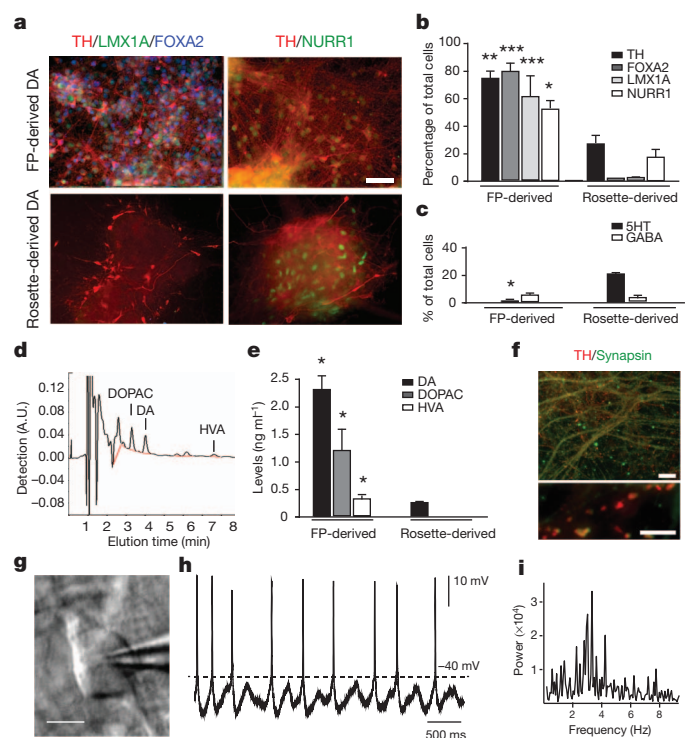


Figure 3 | *In vitro* maturation and functional characterization of FP versus rosette-derived midbrain DA neurons. **a**, Immunocytochemistry at day 50 for TH (red), with LMX1A (green) and FOXA2 (blue; left panels) and NURR1 (green, right panels). **b**, Quantification of TH⁺, FOXA2⁺, LMX1A⁺ and NURR1⁺ cells in rosette- versus FP-derived (LSB/S/F8/CHIR) cultures. **c**, Quantification of serotonin⁺ (5-HT), and GABA⁺ neuronal subtypes at day 50 in rosette- versus FP-derived cultures. **d**, HPLC analysis for DA and metabolites. **e**, Representative HPLC chromatogram in a sample of FP-derived cultures. A.U., arbitrary units. **f**, Levels of DA, DOPAC and HVA in FP- and rosette-derived cultures. **f**, Immunocytochemistry in FP-derived cultures (day 80) for TH (red) and synapsin (green). **g–i**, Electrophysiological analyses of FP cultures at day 80. **g**, Phase contrast image of a patched neuron (**g**) and corresponding recordings (**h**). **i**, Power analysis showing membrane potential oscillations characteristic of DA neuron identity (2–5 Hz). Mean \pm s.e.m.; significance (panels **b**, **c**, **e**) comparing FP versus rosette-derived cultures (Student's *t*-test): ****P* < 0.001; ***P* < 0.01; **P* < 0.05. Scale bars: 50 μ m in **a**, 20 μ m in **f** upper panel and 5 μ m in **f** lower panel, and 20 μ m in **g**.

contaminating cells with proliferative potential. Four and a half months after transplantation FP-derived DA neuron grafts showed a well-defined graft core composed of TH⁺ cells coexpressing *FOXA2* and human-specific NCAM (Fig. 4a–c). Functional analysis showed a complete rescue of amphetamine-induced rotation behaviour. In contrast, rosette-derived grafts had few TH⁺ neurons, did not produce significant reductions in rotation behaviour (Fig. 4d) and showed massive neural overgrowth (graft volume > 20 mm³; Supplementary Fig. 9). Extensive overgrowth reported here as compared to previous work with rosette-derived DA grafts^{27,28} is likely due to the longer survival periods (4.5 months versus 6 weeks), lack of fluorescence-activated cell sorting purification before transplantation and choice of *NOD-SCID IL2Rgc* null host. The number of proliferating Ki-67⁺ cells was minimal in FP-derived grafts (<1% of total cells), while rosette-derived grafts retained pockets of proliferating neural precursors. Neural overgrowth is thought to be caused by primitive anterior neuroectodermal cells within the graft^{6,29}. This hypothesis was supported by the expression of the forebrain marker *FOXG1* in rosette-derived but not FP-derived grafts. A small percentage of astroglial cells were present in both FP- and rosette-derived grafts, although most GFAP⁺ cells were negative for human markers, indicating host origin (Supplementary Fig. 9).

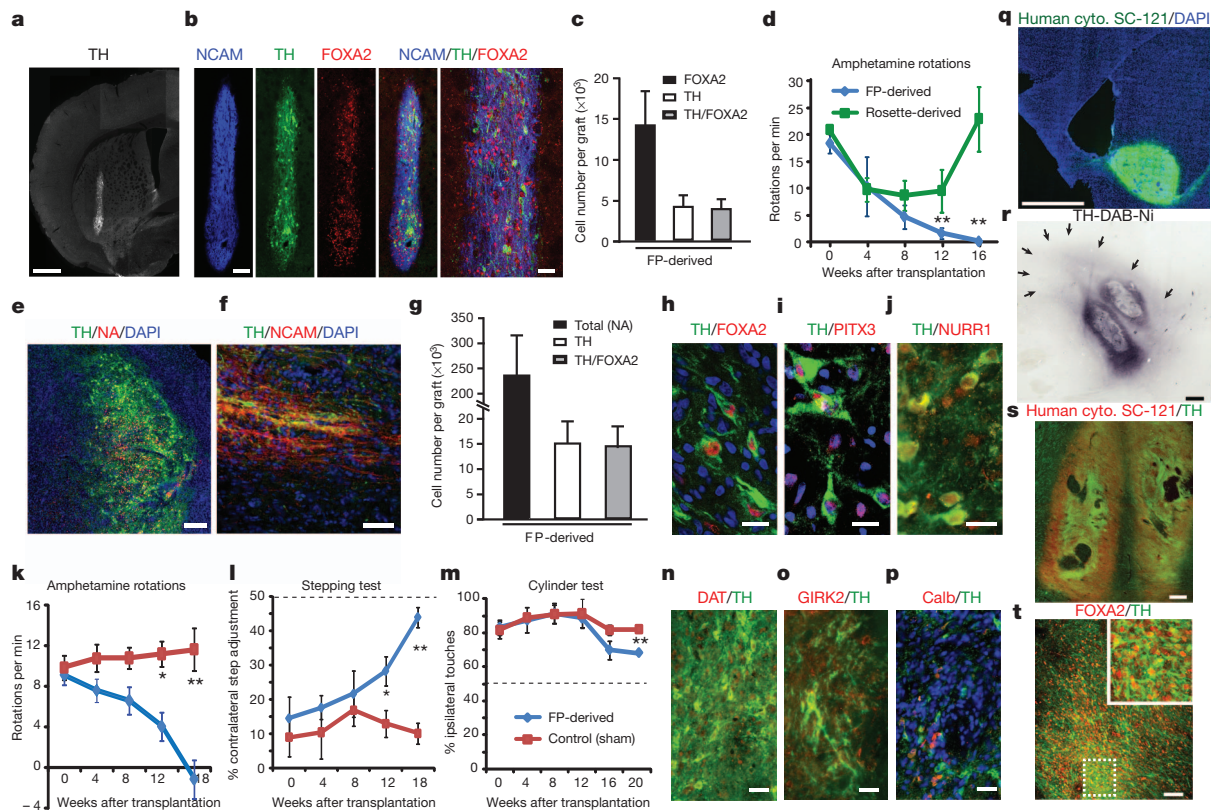


Figure 4 | *In vivo* survival and function of FP-derived human DA neurons in mouse, rat and monkey Parkinson's disease hosts. **a–d**, 6-OHDA-lesioned adult mice (*NOD-SCID IL2Rgc* null strain). **a**, TH expression and graft morphology at 4.5 months after transplantation. **b**, Expression of human-specific marker (NCAM, blue), TH (green) and FOXA2 (red). **c**, Quantification of FOXA2⁺ and TH⁺ cells in FP-derived grafts (mean \pm s.e.m., $n = 4$ at 4.5 months after grafting). **d**, Amphetamine-induced rotation analysis in FP-derived (blue) versus rosette-derived (green) grafts. Scale bars: 500 μ m in **a**, and 100 and 40 μ m in **b** (left and right panels, respectively). **e–p**, 6-OHDA-lesioned adult rats. **e**, **f**, Immunohistochemistry for TH (green) and human-specific markers (red) human nuclear antigen NA (**e**) and NCAM (**f**). **g**, Stereological quantification of NA⁺, TH⁺ and TH⁺ cells co-expressing FOXA2 (average graft volume = 2.6 ± 0.6 mm³). **h–j**, Coexpression of TH (green) with FOXA2 (**h**), PITX3 (**i**) and NURR1 (**j**) (all three red). **k–m**, Behavioural analysis in FP-

versus sham-grafted animals. **k**, Amphetamine-induced rotational asymmetry. **l**, Stepping test: measuring forelimb akinesia in affected versus non-affected side. **m**, Cylinder test: measuring ipsi- versus contra-lateral paw preference. Grafted animals showed significant improvement in all three tests ($P < 0.01$ at 4.5–5 month; $n = 4–6$ each). **n–p**, Immunohistochemistry for TH (green) and coexpression (red) with DAT (**n**), GIRK2 (**o**) and calbindin (**p**). Significance levels (panels **d**, **k**, **l**, **m**): $^{**}P < 0.01$; $^{*}P < 0.05$. Scale bars: 200 μ m in **e**, 50 μ m in **f**, 20 μ m in **h–j** and 40 μ m in **n–p**. **q–t**, Adult MPTP-lesioned rhesus monkeys. **q**, Representative graft at 1 month after transplantation expressing human specific cytoplasm (cyto.) marker SC-121 (green). **r**, TH expression in graft with surrounding TH⁺ fibres (arrows). **s**, Coexpression of SC-121 (red) and TH (green). **t**, Coexpression of FOXA2 (red) and TH⁺ (green). Scale bars: 2 mm for **q**, 500 μ m for **r**, 200 μ m for **s**, and 50 μ m for **t**.

Our results in *NOD-SCID IL2Rgc* null mice demonstrated robust long-term survival of FOXA2⁺/TH⁺ neurons, complete reversal of amphetamine-induced rotation behaviour and lack of neural overgrowth. However, some of these outcomes could be attributable to the specific use of *NOD-SCID IL2Rgc* null mice. To test this hypothesis, FP-derived DA neuron cultures (2.5×10^5 cells) were transplanted in adult 6-OHDA-lesioned rats immunosuppressed pharmacologically using cyclosporine A. Five months after transplantation graft survival was robust (Fig. 4e–h) with an average of more than 15,000 TH⁺ cells coexpressing FOXA2 (Fig. 4g) and human nuclear antigen (Fig. 4e); TH⁺/NCAM⁺ fibres emanated from the graft core into the surrounding host striatum (Fig. 4f). In addition to FOXA2, TH⁺ cells expressed midbrain DA neuron markers PITX3 and NURR1 (Fig. 4h–j). Behavioural analyses showed complete rescue of amphetamine-induced rotational asymmetry in contrast to sham-grafted animals (Fig. 4k). Grafted animals also showed improvements in the stepping test (Fig. 4l) measuring forelimb akinesia and in the cylinder test (Fig. 4m), assays that do not depend on pharmacological stimulation of the DA system. The late onset of recovery (approximately 3–4 months after transplantation) is expected for human DA neurons and depends on the rate of *in vivo* maturation including levels of DAT expression (Fig. 4n). The presence of TH⁺ cells expressing Kir3.2 channels or calbindin indicate that both SNPC (A9) and ventral

tegmental area (A10) DA neurons are present in the graft (Fig. 4o, p). As in mice (Supplementary Fig. 9), serotonergic and GABAergic cells were rare (<1% of total cells), as were the mostly host-derived GFAP⁺ glial cells (7% of total cells; Supplementary Fig. 10). While few serotonin⁺ neurons were detected in the graft, we observed human NCAM-negative, likely host-derived serotonergic fibres (Supplementary Fig. 10). Our results demonstrate excellent graft survival and behavioural outcome in two independent murine models. However, the number of DA neurons required in a mouse or rat brain represents only a fraction of the cells needed in a human. To test the scalability of our approach, we performed pilot grafting studies in two adult 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-lesioned rhesus monkeys. We readily obtained batches of 5×10^7 transplantable DA neuron precursors by day 25 of differentiation using the FP-based protocol. Cells were injected at three locations (posterior caudate and pre-commissural putamen) on each side of the brain (six tracts in total, 1.25×10^6 cells per tract), and the animals were immunosuppressed with cyclosporine-A. One side of the brain was injected with DA precursors from a green fluorescent protein (GFP)-expressing subclone of H9, while the other side was engrafted with cells derived from unmarked H9 cells. One month after transplantation, we observed robust survival of midbrain DA neurons based on expression of GFP (Supplementary Fig. 11) and the human-specific cytoplasmic

marker (SC-121) (Fig. 4q). Each graft core was surrounded by a halo of TH+ fibres extending up to 3 mm into the host (Fig. 4r). The graft cores were composed of TH+ neurons coexpressing SC-121 (Fig. 4s) and FOXA2 (Fig. 4t). Areas within the graft contained Iba1+ host microglia (Supplementary Fig. 11), indicating incomplete immunosuppression.

In conclusion, we present a novel FP-based PSC differentiation protocol that faithfully recapitulates midbrain DA neuron development. Access to cells with the cardinal features of midbrain DA neurons will enable a broad range of biomedical applications such as basic developmental studies, high-throughput drug discovery and Parkinson's disease-iPSC based disease modelling. Importantly, our study establishes a means of obtaining a scalable source of FOXA2+/TH+ neurons for neural transplantation—a major step on the road towards considering a cell based therapy for Parkinson's disease.

METHODS SUMMARY

Human ES cell (H9, H1) and iPSC lines (2C6 and SeV6) were subjected to a modified dual SMAD-inhibition-based¹³ FP induction¹² protocol. Exposure to SHH C25I, purmorphamine, FGF8 and CHIR were optimized for midbrain FP and DA neuron yield (see Fig. 1d). Following FP induction, further maturation was carried out in Neurobasal/B27 medium supplemented with ascorbic acid, BDNF, GDNF, TGFβ3 and dibutyryl (db) cyclic AMP (see full methods for details). The resulting DA neurons were subjected to extensive phenotypic characterization via immunocytochemistry, quantitative PCR with reverse transcription (qRT-PCR), gene expression profiling, high-performance liquid chromatography (HPLC) analysis for DA and *in vitro* electrophysiological recordings. *In vivo* studies were performed in 6-hydroxydopamine-lesioned, hemiparkinsonian rodents (adult *NOD-SCID IL2R γ* mice and Sprague Dawley rats) as well as in two adult rhesus monkeys treated with carotid injections of MPTP. DA neurons were injected stereotactically in the striata of the animals (1.5×10^5 cells in mice, 2.5×10^5 cells in rats) and a total of 7.5×10^6 cells (distributed in six tracts; three on each side of brain) in monkeys. Behavioural assays were performed at monthly intervals after grafting, including amphetamine-mediated rotational analysis as well as a test for focal akinesia ('stepping test') and forelimb use (cylinder test). Rats and mice were killed at 18–20 weeks and the primates at 1 month after grafting. Characterization of the grafts was performed via stereological analyses of cell numbers and graft volumes and comprehensive immunohistochemistry.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 August; accepted 19 October 2011.

Published online 6 November 2011; corrected 21 December 2011 (see full-text HTML version for details).

- Li, X. J. *et al.* Specification of motoneurons from human embryonic stem cells. *Nature Biotechnol.* **23**, 215–221 (2005).
- Perrier, A. L. *et al.* From the cover: derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **101**, 12543–12548 (2004).
- Tabar, V. *et al.* Therapeutic cloning in individual parkinsonian mice. *Nature Med.* **14**, 379–381 (2008).
- Wernig, M. *et al.* Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc. Natl Acad. Sci. USA* **105**, 5856–5861 (2008).
- Lindvall, O. & Kokaia, Z. Stem cells in human neurodegenerative disorders—time for clinical translation? *J. Clin. Invest.* **120**, 29–40 (2010).
- Elkabatz, Y. *et al.* Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22**, 152–165 (2008).
- Roy, N. S. *et al.* Functional engraftment of human ES cell-derived dopaminergic neurons enriched by coculture with telomerase-immortalized midbrain astrocytes. *Nature Med.* **12**, 1259–1268 (2006).
- Kittappa, R., Chang, W. W., Awatramani, R. B. & McKay, R. D. The *foxa2* gene controls the birth and spontaneous degeneration of dopamine neurons in old age. *PLoS Biol.* **5**, e325 (2007).
- Ferri, A. L. *et al.* *Foxa1* and *Foxa2* regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development* **134**, 2761–2769 (2007).
- Roelink, H. *et al.* Floor plate and motor neuron induction by *vhh-1*, a vertebrate homolog of hedgehog expressed by the notochord. *Cell* **76**, 761–775 (1994).

- Liem, K. F., Tremml, G., Roelink, H. & Jessell, T. M. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969–979 (1995).
- Fasano, C. A., Chambers, S. M., Lee, G., Tomishima, M. J. & Studer, L. Efficient derivation of functional floor plate tissue from human embryonic stem cells. *Cell Stem Cell* **6**, 336–347 (2010).
- Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).
- Muroyama, Y., Fujihara, M., Ikeya, M., Kondoh, H. & Takada, S. Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. *Genes Dev.* **16**, 548–553 (2002).
- Joksimovic, M. *et al.* Wnt antagonism of Shh facilitates midbrain floor plate neurogenesis. *Nature Neurosci.* **12**, 125–131 (2009).
- Lyashenko, N. *et al.* Differential requirement for the dual functions of beta-catenin in embryonic stem cell self-renewal and germ layer formation. *Nature Cell Biol.* **13**, 753–761 (2011).
- VanDunk, C., Hunter, L. A. & Gray, P. A. Development, maturation, and necessity of transcription factors in the mouse suprachiasmatic nucleus. *J. Neurosci.* **31**, 6457–6467 (2011).
- Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
- Costa, R. H., Grayson, D. R. & Darnell, J. E., Jr. Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Mol. Cell. Biol.* **9**, 1415–1425 (1989).
- Soldner, F. *et al.* Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964–977 (2009).
- Guzman, J. N., Sanchez-Padilla, J., Chan, C. S. & Surmeier, D. J. Robust pacemaking in substantia nigra dopaminergic neurons. *J. Neurosci.* **29**, 11011–11019 (2009).
- Nedergaard, S., Flatman, J. A. & Engberg, I. Nifedipine- and omega-conotoxin-sensitive Ca^{2+} conductances in guinea-pig substantia nigra pars compacta neurones. *J. Physiol. (Lond.)* **466**, 727–747 (1993).
- Ferrari, D., Sanchez-Pernaute, R., Lee, H., Studer, L. & Isacson, O. Transplanted dopamine neurons derived from primate ES cells preferentially innervate DARPP-32 striatal progenitors within the graft. *Eur. J. Neurosci.* **24**, 1885–1896 (2006).
- Olanow, C. W., Kordower, J. H. & Freeman, T. B. Fetal nigral transplantation as a therapy for Parkinson's disease. *Trends Neurosci.* **19**, 102–109 (1996).
- Zetterstrom, R. H. *et al.* Dopamine neuron agenesis in *Nurr1*-deficient mice. *Science* **276**, 248–250 (1997).
- Quintana, E. *et al.* Efficient tumour formation by single human melanoma cells. *Nature* **456**, 593–598 (2008).
- Kim, H. *et al.* miR-371-3 expression predicts neural differentiation propensity in human pluripotent stem cells. *Cell Stem Cell* **8**, 695–706 (2011).
- Hargus, G. *et al.* Differentiated Parkinson patient-derived induced pluripotent stem cells grow in the adult rodent brain and reduce motor asymmetry in Parkinsonian rats. *Proc. Natl Acad. Sci. USA* **107**, 15921–15926 (2010).
- Aubry, L. *et al.* Striatal progenitors derived from human ES cells mature into DARPP32 neurons *in vitro* and in quinolinic acid-lesioned rats. *Proc. Natl Acad. Sci. USA* **105**, 16707–16712 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Manova, M. Tomishima and A. Viale for excellent technical support, and R. McKay for the anti-nestin antibody. The work was supported by NIH/NINDS grant NS052671, the European Commission project NeuroStemcell, the Starr foundation and NYSYSTEM contract C024414 to L.S.; by NYSYSTEM contract C024413, the Michael T. McCarthy Foundation and the Elkus Family Foundation to V.T.; by the Consolidated Anti-Aging Foundation to J.H.K.; by NIH/NINDS grant P50 NS047085, and support from Falk Medical Research Trust to D.J.S.; J.-W.S. was supported by NYSCF (Druckenmiller fellowship) and S.K. by a Starr stem cell scholar fellowship.

Author Contributions S.K. and J.-W.S.: conception and study design, maintenance and directed differentiation of PSCs, cellular/molecular assays, histological analyses, mouse behavioural assays, data interpretation and writing of manuscript. J.P., G.A., C.A. and A.B.: rat transplantation, histological analyses and behavioural assays. Y.M.G.: mouse transplantation and histological analyses. D.R.W. and J.H.K.: monkey transplantation, histological analysis and data interpretation. L.Y. and M.F.B.: HPLC analysis and data interpretation. L.C.-R., Z.X. and D.J.S.: electrophysiological analyses and data interpretation. V.T.: study design, data analysis and writing of manuscript. L.S.: conception and study design, data analysis and interpretation, and writing of manuscript.

Author Information The raw gene expression data generated in this study have been deposited in the Gene Expression Omnibus database under accession number GSE32658. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.S. (studerl@mskcc.org).

METHODS

Culture of undifferentiated ES cells. ES cell lines H9 (WA-09, XX, passages 35–45), H1 (WA-01, XY, passages 30–40) and iPS cell lines 2C6 (XY, passages 20–30, ref. 27) and SeV6 (XY, passages 20–30; derived from MRC-5 using non-integrating 4 factor Sendai vector system³⁰) were maintained on mouse embryonic fibroblasts (MEF, Global Stem) in 20% knockout serum replacement (KSR, Invitrogen)-containing ES cell medium as described previously²⁷.

Neural induction. For FP-based midbrain DA neuron induction, a modified version of the dual-SMAD inhibition¹³ and FP induction¹² protocol was used based on timed exposure to LDN193189 (100 nM, Stemgent), SB431542 (10 μ M, Tocris), SHH C25II (100 ng ml⁻¹, R&D), Purmorphamine (2 μ M, Stemgent), FGF8 (100 ng ml⁻¹, R&D) and CHIR99021 (CHIR; 3 μ M, Stemgent). Note: for the FP induction protocol we refer to 'SHH' treatment as exposure of cells to a combination of SHH C25II 100 ng ml⁻¹ + purmorphamine (2 μ M). Cells were plated (35 \times 10³–40 \times 10³ cells per cm²) and grown for 11 days on matrigel (BD) in knockout serum replacement medium (KSR) containing DMEM, 15% knockout serum replacement, 2 mM L-glutamine and 10 μ M β -mercaptoethanol. KSR medium was gradually shifted to N2 medium starting on day 5 of differentiation as described previously¹³. On day 11, media was changed to Neurobasal/B27/L-Glut containing medium (NB/B27; Invitrogen) supplemented with CHIR (until day 13) and with BDNF (brain-derived neurotrophic factor, 20 ng ml⁻¹; R&D), ascorbic acid (0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng ml⁻¹; R&D), TGF β 3 (transforming growth factor type β 3, 1 ng ml⁻¹; R&D), dibutyl cAMP (0.5 mM; Sigma), and DAPT (10 μ M; Tocris), for 9 days. On day 20, cells were dissociated using Accutase (Innovative Cell Technology) and replated under high cell density conditions (300 \times 10³–400 \times 10³ cells per cm²) on dishes pre-coated with polyornithine (PO; 15 μ g ml⁻¹)/laminin (1 μ g ml⁻¹)/fibronectin (2 μ g ml⁻¹) in differentiation medium (NB/B27 + BDNF, ascorbic acid, GDNF, dbcAMP, TGF β 3 and DAPT) until the desired maturation stage for a given experiment.

For rosette-based DA neuron induction we followed our previously described protocols² but used dual-SMAD inhibition to accelerate the initial neural induction step. In brief, ES cells were induced towards neural fate by co-culture with irradiated M55 cells in KSR supplemented with SB431542 and Noggin (250 ng ml⁻¹; R&D), from day 2–8 and SHH + FGF8 from day 6–11 of differentiation. After 11 days in KSR, neural rosettes were manually isolated and cultured (P1 stage) in N2 medium supplemented with SHH, FGF8, BDNF and ascorbic acid as described previously². After 5–7 days in P1 stage, rosettes were again harvested mechanically and triturated following incubation in Ca²⁺/Mg²⁺-free Hanks' balanced salt solution (HBSS) for 1 h and replated on PO/laminin/fibronectin-coated plates. Patterning with SHH/FGF8 was continued for 7 days at P2 stage followed by final differentiation in the presence of BDNF, ascorbic acid, GDNF, TGF β 3 and dbcAMP as described above until the desired maturation stage for a given experiment (typically 5–7 days for transplantation studies or 32 days for *in vitro* functional studies).

Gene expression analyses. Total RNA was extracted during differentiation at days 0, 1, 3, 5, 7, 9, 11, 13 and 25 from each condition of control LSB, LSB/SHH/FGF8 and LSB/SHH/FGF8/CHIR using an RNeasy kit (Qiagen). For microarray analysis, total RNA was processed by the MSKCC Genomic core facility and hybridized on Illumina Human ref-12 bead arrays according to the specifications of the manufacturer. Comparisons were performed among each days and conditions using the LIMMA package from Bioconductor (<http://www.bioconductor.org>). Genes found to have an adjusted *P*-value <0.05 and a fold change greater than two were considered significant. Some of the descriptive microarray data analyses and presentation was performed using a commercially available software package (Partek Genomics Suite (version 6.10.0915)). For qRT-PCR analyses, total RNA at day 25 of each condition was reverse transcribed (Quantitech, Qiagen) and amplified material was detected using commercially available TaqMan gene expression assays (Applied Biosystems) with the data normalized to HPRT. Each data point represents nine technical replicates from three independent biological samples. Raw data of all the microarray studies are available at GEO <http://www.ncbi.nlm.nih.gov/geo/> accession number GSE32658.

Animal surgery. All rodent and monkey procedures were performed following NIH guidelines, and were approved by the local Institutional Animal Care and Use Committee (IACUC), the Institutional Biosafety Committee (IBC) as well as the Embryonic Stem Cell Research Committee (ESCRO).

Mice: *NOD-SCID IL2R γ* null mice (20–35 g; Jackson Laboratory) were anaesthetized with ketamine (90 mg kg⁻¹; Akorn) and xylazine (4 mg kg⁻¹). 6-hydroxydopamine (10 μ g 6-OHDA, Sigma-Aldrich) was injected stereotactically into the striatum at the following coordinates (in millimetres): AP, 0.5 (from bregma); ML, -2.0; DV, -3.0 (from dura). Mice with successful lesions (an average of >6 rotations per minute) were selected for transplantation. A total of 1.5 \times 10⁵ cells were injected in a volume of 1.5 μ l into the striatum at the following

coordinates (in mm): AP, 0.5; ML, -1.8; DV, 3.2. The mice were killed 18 weeks after transplantation.

Rats: adult female Sprague Dawley (Taconic) rats (180–230 g) were anaesthetized with Ketamine (90 mg kg⁻¹) and xylazine (4 mg kg⁻¹) during surgical procedures. Unilateral, medial forebrain bundle lesions of the nigro-striatal pathway were established by stereotaxic injection of 6-OHDA (3.6 mg ml⁻¹ in 0.2% ascorbic acid and 0.9% saline, Sigma) at two sites³¹. Rats were selected for transplantation if amphetamine-induced rotation exceeded 6 rotations per min by 6–8 weeks post injection. 2.5 \times 10⁵ cells were transplanted into the striatum of each animal (Coordinates: AP 1.0 mm, ML -2.5 mm and V -4.7 mm; toothbar set at -2.5). Control rats received PBS instead. The surgical procedures were described previously³¹. Daily intraperitoneal injections of cyclosporine 15 mg kg⁻¹ (Bedford Labs) were started 24 h before cell grafting and continued until death, 20 weeks following cell grafting.

Primates: two adult (17–18 year old; 10–12 kg; female) rhesus monkeys were rendered hemiparkinsonian via carotid MPTP administration followed by weekly intravenous MPTP administration to create a bilateral parkinsonian syndrome³². Both animals displayed parkinsonian symptoms consistent with a moderately-severe lesion. On the day of transplantation surgery, animals were tranquilized with ketamine (3.0 mg kg⁻¹, intramuscular) and dexdomitor (0.02–0.04 mg kg⁻¹ intramuscular), intubated to maintain a stable airway and anaesthetized with isoflurane. They were then placed into a stereotaxic frame for surgery. Both rhesus monkeys underwent a single surgery with three intracranial injections of human FP-derived DA cultures based on stereotaxic coordinates³³. Bilateral injections of cells (10 μ l/injection; 125,000 cells per μ l) were performed at three sites (one posterior caudate, two pre-commissural putamen and overlying white matter) for a total volume of 30 μ l per hemisphere. An infusion pump attached to a stereotaxic micromanipulator was used to deliver the cells at a rate of 1 μ l min⁻¹ though a 50- μ l Hamilton syringe with 28 G needle. After the injections were completed, the needle was left in place for an additional 2–5 min to allow the infusate to diffuse off the needle tip before slowly retracting the syringe. Immediately following surgery, the animals received analgesics (buprenex, 0.01 mg kg⁻¹ intramuscular, BID for 72 h post surgery; meloxicam, 0.1 mg kg⁻¹ subcutaneous, SID for 72 h post surgery) as well as an antibiotic (cephazolin, 25 mg kg⁻¹ intramuscular, BID) until 72 h post-surgery. The animals received cyclosporine A (Neoral, Sandimmune) orally (30 mg kg⁻¹ tapered to 15 mg kg⁻¹) once daily beginning 48 h before surgery until death, one month following transplantation.

Behavioural assays. Amphetamine-induced rotations (mice and rats) and the stepping test (rat) were carried out before transplantation and 4, 8, 12 and 18 weeks after transplantation. Rotation behaviour in mice was recorded 10 min after intraperitoneal injection of d-amphetamine (10 mg kg⁻¹, Sigma) and recorded for 30 min. Rotation behaviour in rats was recorded 40 min after intraperitoneal injection of d-amphetamine (5 mg kg⁻¹) and automatically assessed by the TSE VideoMot2 system (Germany). The data were presented as the average number of rotations per minute. The stepping test was modified from^{34,35}. In brief, each rat was placed on a flat surface, its hind legs were lifted by gently holding up the tail to allow only the forepaws to touch the table. The experimenter pulled the rat backwards 1 m at a steady pace. Adjusting step numbers from both contralateral and ipsilateral forepaws were counted. Data was presented as the percentage of contralateral/(contralateral + ipsilateral) adjusting steps. The cylinder test was performed by placing each animal in a glass cylinder and counting the number of ipsilateral versus contralateral paw touches (out of 20 touches) to the wall of the cylinder as described previously³.

Tissue processing. Mice and rats received overdoses of pentobarbital intraperitoneally (50 mg kg⁻¹) to induce deep anaesthesia and were perfused in 4% paraformaldehyde (PFA). Brains were extracted, post-fixed in 4% PFA then soaked in 30% sucrose solutions for 2–5 days. They were sectioned on a cryostat after embedding in O.C.T. (Sakura-Finetek).

Primates were killed under deep anaesthesia with ketamine (10 mg kg⁻¹, intramuscular) and pentobarbital (25 mg kg⁻¹, intravenous) via cardiac perfusion with heparinized 0.9% saline followed by fresh cold 4% PFA fixative (pH 7.4). Immediately following primary fixation, brains were removed from the skull and post-fixed in 4% PFA, free-floating, for 24–36 h. They were then rinsed and re-suspended in 10% sucrose on a slow shaker at 4 °C, and allowed to 'sink'. The process was then repeated in 20% sucrose followed by 30% sucrose. Whole brains were cut coronally into 40 μ m serial sections on a frozen sledge microtome and stored free-floating in cryopreservative medium at -20 °C.

Immunohistochemistry. Cells were fixed in 4% PFA and blocked with 1% BSA with 0.3% Triton. Brain tissue sections were washed in cold PBS and processed similarly. Primary antibodies were diluted in 1–5% BSA or Normal Goat Serum and incubated according to manufacturer recommendations. A comprehensive list of antibodies and sources is provided as Supplementary Table 6. Appropriate

Alexa488, Alexa555 or Alexa647-conjugated secondary antibodies (Molecular Probes) were used with 4',6-diamidino-2-phenylindole (DAPI) nuclear counterstain (Thermo Fisher). For some analyses biotinylated secondary antibodies were used followed by visualization via DAB chromogen.

HPLC analysis. Reversed-phase HPLC with electrochemical detection for measuring levels of DA, HVA and DOPAC was performed as described previously^{7,36}. Culture samples were collected in perchloric acid at day 65 of differentiation. For some experiments DA was measured directly in the medium using the same detection system but following aluminium extraction of DA and its metabolites using a commercially available kit as described previously³⁶.

Electrophysiological recordings. Cultures were transferred to a recording chamber on an upright microscope equipped with a $\times 40$ water-immersion objective (Eclipse E600FN; Nikon); cultures were perfused with saline containing in mM: 125 NaCl, 2.5 KCl, 25 NaHCO₃, 1.25 NaH₂PO₄, 2 CaCl₂, 1 MgCl₂ and 25 glucose (34 °C; saturated with 95% O₂/5% CO₂; pH 7.4; 298 mOsm l⁻¹). The saline flow rate was 2–3 ml min⁻¹ running through an in-line heater (SH-27B with TC-324B controller; Warner Instruments). Neurons were visualized by video microscopy with a cooled-CCD digital camera (CoolSNAP ES², Photometrics, Roper Scientific). Cells selected for electrophysiological recordings had neuron-like shapes with fine branching neurites. Somatic whole-cell patch-clamp recordings in current clamp configuration were performed with a MultiClamp 700B amplifier (Molecular Devices). Signals were filtered at 1–4 kHz and digitized at 5–20 kHz with a Digidata 1440A (Molecular Devices). Recording patch electrodes were fabricated from filamented borosilicate glass (Sutter Instruments) pulled on a Flaming-Brown puller (P-97, Sutter Instruments) and had resistances of 4–6 M Ω in the bath. Electrodes were filled with internal solution containing in mM: 135 K-MeSO₄, 5 KCl, 5 HEPES, 0.25 EGTA, 10 phosphocreatine-di(tris), 2 ATP-Mg and 0.5 GTP-Na (pH 7.3, osmolarity adjusted to 290–300 mOsm l⁻¹). The amplifier bridge circuit was adjusted to compensate for electrode resistance and monitored. Electrode capacitance was also compensated. If series resistance increased >20% during the recording, the data were discarded.

Cell counts and stereological analyses. The percentages of marker positive cells at the FP (day 11), midbrain DA neuron precursor (day 25) and mature DA

neuron stages (day 50 or later) were determined in samples derived from at least three independent experiments each. Images for quantification were selected in a uniform random manner and each image was scored first for the number of DAPI-positive nuclei, followed by counting the number of cells expressing the marker of interest. All data are presented as mean \pm s.e.m. Quantification of human cells (identified with anti-human nuclear antigen) and TH+ neurons within grafts was performed on every tenth section where a graft was identifiable. Cell counts and graft volume were determined using the optical fractionator probe and the Cavalieri estimator using the Stereo Investigator software (MBF bioscience, Vermont) as described previously³⁷. Data are presented as estimated total cell number and total graft volume \pm s.e.m. Statistical analysis was performed using the Student *t*-test (comparing two groups) or ANOVA with Dunnett test (comparing multiple groups against control).

30. Ban, H. *et al.* Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *Proc. Natl. Acad. Sci. USA* **108**, 14234–14239 (2011).
31. Studer, L., Tabar, V. & McKay, R. D. Transplantation of expanded mesencephalic precursors leads to recovery in parkinsonian rats. *Nature Neurosci.* **1**, 290–295 (1998).
32. Kordower, J. H. *et al.* Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease. *Science* **290**, 767–773 (2000).
33. Paxinos, G., Huang, X.-F. & Toga, A. W. *The Rhesus Monkey Brain in Stereotaxic Coordinates* (Academic Press, 2000).
34. Blume, S. R., Cass, D. K. & Tseng, K. Y. Stepping test in mice: a reliable approach in determining forelimb akinesia in MPTP-induced Parkinsonism. *Exp. Neurol.* **219**, 208–211 (2009).
35. Crawley, J. N. *What's Wrong With My Mouse: Behavioral Phenotyping of Transgenic and Knockout Mice* (Wiley, 2000).
36. Studer, L. *et al.* Noninvasive dopamine determination by reversed phase HPLC in the medium of free-floating roller tube cultures of rat fetal ventral mesencephalon: A tool to assess dopaminergic tissue prior to grafting. *Brain Res. Bull.* **41**, 143–150 (1996).
37. Tabar, V. *et al.* Migration and differentiation of neural precursors derived from human embryonic stem cells in the rat brain. *Nature Biotechnol.* **23**, 601–606 (2005).

Cryptochromes mediate rhythmic repression of the glucocorticoid receptor

Katja A. Lamia^{1,2}, Stephanie J. Papp², Ruth T. Yu¹, Grant D. Barish¹, N. Henriette Uhlenhaut¹†, Johan W. Jonker¹†, Michael Downes¹ & Ronald M. Evans¹

Mammalian metabolism is highly circadian and major hormonal circuits involving nuclear hormone receptors display interlinked diurnal cycling^{1,2}. However, mechanisms that logically explain the coordination of nuclear hormone receptors and the clock are poorly understood. Here we show that two circadian co-regulators, cryptochromes 1 and 2, interact with the glucocorticoid receptor in a ligand-dependent fashion and globally alter the transcriptional response to glucocorticoids in mouse embryonic fibroblasts: cryptochrome deficiency vastly decreases gene repression and approximately doubles the number of dexamethasone-induced genes, suggesting that cryptochromes broadly oppose glucocorticoid receptor activation and promote repression. In mice, genetic loss of cryptochrome 1 and/or 2 results in glucose intolerance and constitutively high levels of circulating corticosterone, suggesting reduced suppression of the hypothalamic–pituitary–adrenal axis coupled with increased glucocorticoid transactivation in the liver. Genomically, cryptochromes 1 and 2 associate with a glucocorticoid response element in the phosphoenolpyruvate carboxykinase 1 promoter in a hormone-dependent manner, and dexamethasone-induced transcription of the phosphoenolpyruvate carboxykinase 1 gene was strikingly increased in cryptochrome-deficient livers. These results reveal a specific mechanism through which cryptochromes couple the activity of clock and receptor target genes to complex genomic circuits underpinning normal metabolic homeostasis.

Glucocorticoids (cortisol in humans and corticosterone in rodents) are critical regulators of many aspects of mammalian physiology, including glucose homeostasis and immune function. These stress hormones exhibit a robust diurnal rhythm in the circulation. Synthetic glucocorticoids are widely used as anti-inflammatory drugs, but cause undesirable side effects including hyperglycaemia, insulin resistance and suppression of adrenal function. A paradox of nuclear receptor signalling is that potent agonists typically repress as many genes as they activate. The cloning of the glucocorticoid receptor³ revealed that the same gene product mediated both agonist-induced activation and repression through dichotomous mechanisms that remain poorly understood. Ligand stimulation results in translocation of glucocorticoid receptor from the cytoplasm to the nucleus where it binds to glucocorticoid response elements (GREs) and alters the transcription of hundreds to thousands of genes. Several cofactors have been identified that mediate transcriptional activation, including the steroid receptor coactivators (SRC1–3), the histone acetyltransferases CBP/p300 and the nuclear methylase coactivator-associated arginine methyltransferase 1 (CARM1)⁴. For glucocorticoid-receptor-dependent repression, combinations of direct and indirect mechanisms have been proposed. For example, glucocorticoid receptor may ‘tether’ or bind to pro-inflammatory transcription factors, such as nuclear factor of kappa light polypeptide gene enhancer in B cells (NF- κ B), and ‘trans-repress’ gene expression by interfering with their activation mechanism⁵. In addition, glucocorticoids can directly repress metabolic

targets, particularly in the hypothalamic–pituitary–adrenal (HPA) axis, in a way that does not seem to require NF- κ B or other inflammatory transcription factors. In this case, glucocorticoid receptor acts through direct binding to negative response elements that presumably promote recruitment of one or more hypothetical repressive cofactors.

Circadian clocks drive rhythms in physiology and behaviour that enable organisms to keep track of the time of day and to adjust their physiology to adapt to recurrent, and therefore predictable, changes in the external environment. Mammalian circadian clocks are based on a transcription and translation feedback loop. In this, a heterodimer of the transcription factors circadian locomotor output cycles kaput (CLOCK) and brain and muscle ARNT-like 1 (BMAL1) drives transcription from E-box elements, including that of their own repressors, the period (*Per1*, *Per2* and *Per3*) and cryptochrome (*Cry1* and *Cry2*) genes². For many years, mammalian clocks were thought to reside solely in the hypothalamic suprachiasmatic nucleus, in which the clock is set by light signals through the retinohypothalamic tract, and from which secreted signals drive rhythms in locomotor activity. Subsequently, circadian clocks have been identified throughout mammalian peripheral organs and their timing demonstrated to be determined by metabolic cues rather than in direct response to light⁶. Several molecular mechanisms have been described that contribute to metabolic resetting of clock time in peripheral organs^{7–9}.

The existence of clocks in peripheral organs and the demonstration that their timing is set by metabolic cues led to the question of whether they participate in physiological regulation that enables organisms to adapt to changing metabolic needs over the course of the day. Tissue-specific ablation of clock function demonstrated that clocks in the liver¹⁰ and pancreas^{11,12} indeed contribute to glucose homeostasis at specific times of the day. Several studies have demonstrated that thousands of transcripts are subject to circadian regulation in myriad organs^{13,14}, and that most of this regulation is lost upon tissue-specific clock disruption^{10,15}, suggesting that rhythmic transcription of rate-limiting enzymes and transporters contributes to circadian physiological regulation¹⁴.

The time of peak expression for these oscillating transcripts is distributed across the day, suggesting that multiple mechanisms contribute to their diurnal regulation. In addition to cascades of rhythmically expressed transcription factors, including nuclear hormone receptors, which can be driven by clock-dependent rhythmic expression, circadian repression of subsets of target genes by the negative arm of the clock may contribute to oscillation of specific genetic programs. The direct regulation of cryptochromes by metabolic signals⁷ makes them ideal potential crossover regulators for circadian and metabolic gene expression programs. Because nuclear receptors are prototypic regulators of metabolism, we examined the possibility that cryptochromes 1 (*Cry1*) and 2 (*Cry2*) might contribute to diurnal aspects of their regulatory output¹.

We first explored whether *Cry1* can interact with two nuclear hormone receptors known to modulate clock function, the retinoid-related orphan receptors (ROR- $\alpha/\beta/\gamma$) and Rev-Erb- α/β . Unexpectedly,

¹Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ²Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. †Present addresses: Max-Delbrueck-Centrum fuer Molekulare Medizin, Genetics and Physiology, Berlin, 13125, Germany (N.H.U.); Center for Liver, Digestive and Metabolic Diseases, University Medical Center Groningen, The Netherlands (J.W.J.).

we find robust association between Cry1 and ROR- α and γ but not with ROR- β or Rev-Erb- α/β though we did reproduce the reported association of PER2 with Rev-Erb- α ¹⁶ (Fig. 1a, b). This is intriguing for two reasons: (1) ROR- α and ROR- γ but not ROR- β can activate Bmal1 transcription¹⁷ and (2) Cry1 preferentially associates with the activation (ROR) and not the repressive (Rev-Erb) arm of the pathway. Together, these results suggest that the so-called 'secondary loop' in which RORs and Rev-Erbs modulate BMAL1 transcription and are thought to make clocks more robust and accurate involves cryptochromes.

We next initiated a broader survey of nuclear hormone receptor interactions and unexpectedly discovered a robust physical association of Cry1 with the glucocorticoid and androgen receptors (Fig. 1a). Although Per2 can interact with some nuclear hormone receptors (such as Rev-Erb¹⁶), it does not appear to bind glucocorticoid receptor (Fig. 1c). Further analysis reveals that Cry2 also interacts with glucocorticoid receptor (Fig. 1d) and that the association is stimulated by glucocorticoids (Fig. 1e, f). Notably, Cry1 represses the ability of glucocorticoid receptor to drive expression of a luciferase reporter from a GRE-containing promoter (Fig. 1g). Finally, we used a series of deletion mutants of glucocorticoid receptor to demonstrate that Cry1 interacts with the carboxy (C) terminus of glucocorticoid receptor (Fig. 1h), which is required for either activation or repression of transcription in response to ligand. Steroid hormones (glucocorticoids, testosterone, progesterone and aldosterone) are critical regulators of metabolic and reproductive physiology, so the possibility that they could be directly regulated by cryptochromes suggests a new mechanism by which circadian clocks modulate physiological rhythms.

To determine whether cryptochromes modulate glucocorticoid-stimulated changes in transcription of endogenous genes *in vivo*, we analysed global changes in gene expression after glucocorticoid treatment of primary fibroblasts derived from wild-type and *cry1*^{-/-}; *cry2*^{-/-} double knockout (DKO) littermate mouse embryos

(Fig. 2a). Circadian rhythms of transcription can be synchronized by dexamethasone treatment of fibroblasts¹⁸. These changes are expected to be lost in Cry-deficient cells, such that a subset of both positive and negative transcriptional responses to dexamethasone will be absent. However, if cryptochromes are independently capable of repressing transcription through glucocorticoid receptor, we would expect to see a global positive shift in the transcriptional response to dexamethasone (for example, an increase in gene activation and a decrease in repression). Indeed, we found that the transcriptional response to the synthetic glucocorticoid dexamethasone was significantly more positive in cryptochrome-deficient cells: 205 more genes (191→396) were activated by dexamethasone in Cry-deficient cells than in controls, indicating that Cry limits induction of a larger network of glucocorticoid receptor responsive genes. Perhaps even more unexpectedly, of 657 dexamethasone-repressed genes only 64 were suppressed in DKO cells (Fig. 2b). We confirmed the observed expression patterns for several individual genes in an independently derived pair of control and cryptochrome-deficient mouse embryonic fibroblasts (MEFs) (Fig. 2c). Because the interpretation of the results for individual transcripts is complicated by the potential role of circadian synchronization, we also examined the expression profile of the established glucocorticoid receptor transcriptional target serum/glucocorticoid regulated kinase 1 (*sgk1*) in response to dexamethasone stimulation. We found that *sgk1* messenger RNA (mRNA) is more robustly activated in *cry1*^{-/-}; *cry2*^{-/-} cells than control cells within 2 h of exposure to glucocorticoids (Fig. 2d) and over a range of dosages (Fig. 2e). In general, the dynamic response to dexamethasone is complex and transcript dependent but more positive in the absence of cryptochromes (Supplementary Fig. 2). Together, these results lead to the surprising suggestion that Cry1 and Cry2, central regulators of the circadian clock, also directly oppose dexamethasone-induced glucocorticoid receptor activation.

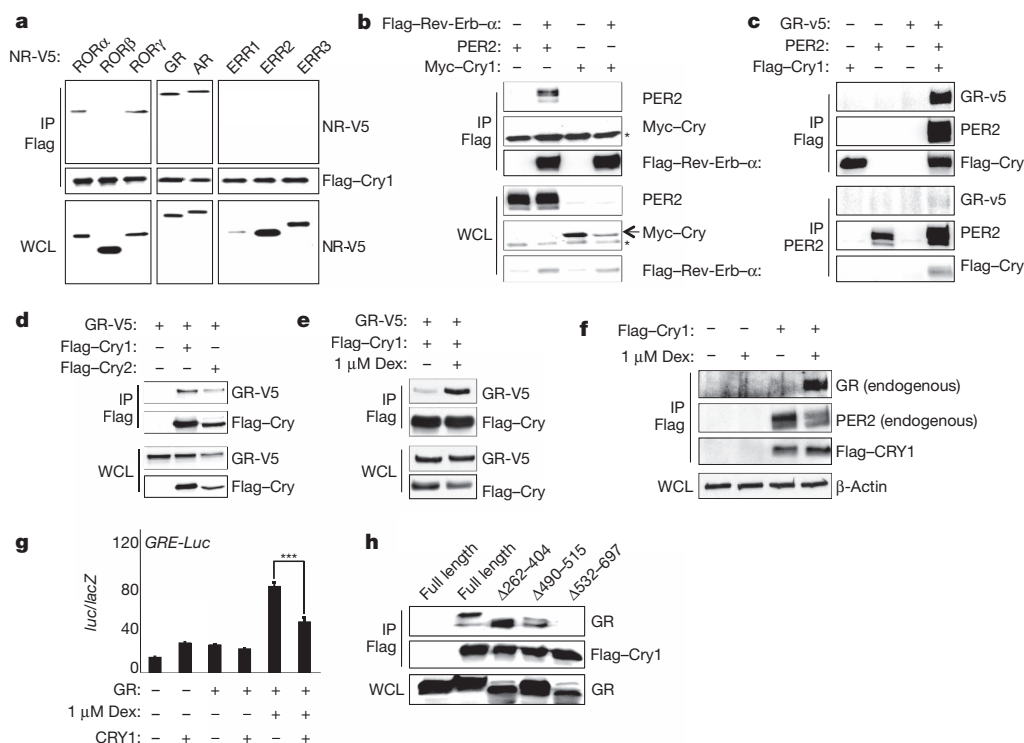


Figure 1 | Cryptochromes interact with glucocorticoid receptor.

a–e, h, Immunoblots showing recovery of the indicated proteins from 293T cells expressing the indicated plasmids after immunoprecipitation with the indicated antibodies. In **e**, cells were treated with vehicle (–) or 1 μ M dexamethasone (+) for 16 h. NR, nuclear hormone receptor; GR, glucocorticoid receptor; AR, androgen receptor; WCL, whole-cell lysate; IP;

immunoprecipitated. **f**, Immunoblots showing recovery of endogenous glucocorticoid receptor and PER2 from MEFs stably expressing empty vector (–) or Flag-Cry1 (+) after Flag immunoprecipitation. **g**, Luciferase activity in CV-1 cells transfected as indicated. Data represent the mean \pm s.e.m. of triplicate samples. *** P < 0.001. Data represent typical results of two to six independent experiments.

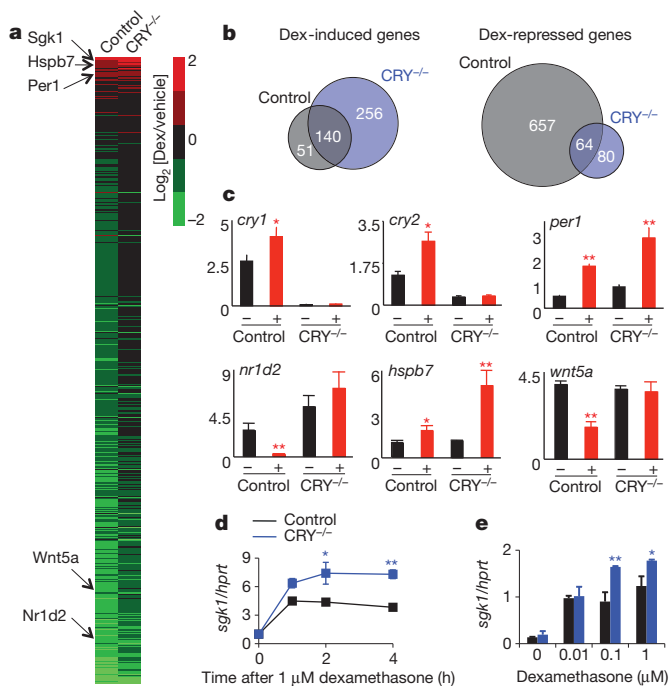


Figure 2 | Cryptochromes modulate glucocorticoid-receptor-dependent transcription. **a**, Heat map: colour denotes dexamethasone (Dex)-induced change in expression of all transcripts significantly altered by genotype or glucocorticoid stimulation. Selected genes are indicated on the left. **b**, Transcripts altered by dexamethasone treatment in control (grey circles) or DKO (blue circles) MEFs. **c**, Expression of indicated transcripts in control or DKO MEFs after overnight treatment with vehicle (-) or dexamethasone (+). **d**, **e**, Expression of *sgk1* in MEFs after treatment with dexamethasone for 1–4 h. In **c–e**, data represent the mean ± s.e.m. of triplicate samples analysed in triplicate. **P* < 0.01, ***P* < 0.001.

Some of the best-studied physiological roles of glucocorticoids are suppression of the immune system and activation of gluconeogenesis. We examined whether cryptochromes are required for transrepression of inflammatory genes in primary bone marrow macrophages from wild-type and DKO mice. After lipopolysaccharide stimulation with or without previous dexamethasone treatment, the expression of tumour-necrosis factor alpha (TNF-α) and chemokine (C-C motif) ligand 4 (Ccl4) was indistinguishable in wild-type and DKO macrophages (Supplementary Fig. 3). The basal and lipopolysaccharide-induced expression of monocyte chemoattractant protein 1, interleukin 6 and inducible nitric oxide synthase was elevated in macrophages lacking Cry1 and Cry2, but these transcripts were equally suppressed by dexamethasone treatment regardless of the presence or absence of cryptochromes (Supplementary Fig. 3). Thus, cryptochromes appear to regulate a distinct, albeit substantial, subset of glucocorticoid receptor regulated target genes, which does not include the NF-κB inflammatory gene network.

Chronic treatment with glucocorticoids to suppress inflammation often results in hyperglycaemia due to glucocorticoid-receptor-induced expression of phosphoenolpyruvate carboxykinase 1 (Pck1), a rate-limiting gluconeogenic enzyme in the liver. If cryptochrome repression of glucocorticoid-receptor-mediated transcription were relevant in the liver, we would expect the ability of glucocorticoids to induce *pck1* transcription to be lower when cryptochromes are present. Indeed, *pck1* induction after a 1 h exposure to dexamethasone was dependent on the time of day and the induction of *pck1* is inversely correlated with the amount of cryptochrome protein in the nucleus (Fig. 3a). Excluding the possibility that the observed effect of day-time involves altered expression or chromatin accessibility of glucocorticoid receptor, in an independent experiment, we observed a greater increase in *pck1* expression in response to dexamethasone at ZT4

(zeitgeber time, which denotes hours after the lights are turned on) than at ZT16, even though the glucocorticoid-induced association of glucocorticoid receptor with the Pck1 promoter was relatively elevated at ZT16 (Supplementary Fig. 4). Furthermore, using antibodies that we generated to detect endogenous Cry1 and Cry2 (Supplementary Fig. 5), we found that cryptochromes interact with glucocorticoid receptor in the liver after ligand stimulation *in vivo* at night (ZT18) when glucocorticoids are less effective at inducing the expression of *pck1* (Fig. 3b). (*Cry1*^{-/-} and *Cry2*^{-/-} samples were used in this experiment to ensure antibody specificity without abolishing circadian rhythmicity.) Strikingly, dexamethasone stimulated *pck1* transcription 12.8-fold at ZT16 in Cry-deficient livers, even though glucocorticoid-induced association of glucocorticoid receptor with the Pck1 promoter was reduced in the DKO livers (Fig. 3c). Finally, chromatin immunoprecipitation of Cry1 and Cry2 revealed that cryptochromes are associated with the Pck1 promoter GRE after treatment with dexamethasone *in vivo* in mouse livers, whereas dexamethasone had no effect on the Bmal1-mediated association of Cry1 and Cry2 with the Dbp promoter (Fig. 3d). Together, these data demonstrate that cryptochromes directly repress glucocorticoid receptor induction of *pck1* in the liver and thus may limit glucocorticoid-induced hyperglycaemia.

To examine the role of cryptochromes in glucocorticoid-dependent physiology, we measured several glucocorticoid-receptor-dependent parameters in wild-type and Cry-deficient mice. (Importantly, Cry-deficient mice were maintained in 12:12 light:dark conditions and were found to have normal behavioural patterns, Supplementary Fig. 6.) In addition to suppressing inflammation and increasing gluconeogenesis through *pck1*, a principal role of glucocorticoids is to promote negative feedback in all three organs of the (HPA) axis¹⁹. However, in DKO mice, we find that glucocorticoids, which normally show robust cycling rhythm, fail to be suppressed during the day as they are in wild-type mice (Fig. 4a). Perhaps not surprisingly, glucose homeostasis is severely disrupted in Cry-deficient mice, consistent with the observed regulation of *pck1* transcription: thus, DKO mice exhibit elevated blood glucose in response to acute feeding after an overnight fast (Fig. 4b) and severely impaired glucose clearance in a glucose tolerance test (Fig. 4c). Even mice lacking either Cry1 (*cry1*^{-/-}) or Cry2 (*cry2*^{-/-}) were significantly impaired in their ability to restore normal blood glucose after glucose injection (Fig. 4c). In contrast, Cry-deficient animals were normally responsive to insulin (Fig. 4d).

To examine directly the role of cryptochromes in repression of the HPA axis and in glucocorticoid-induced hyperglycaemia, we subjected wild-type and littermate DKO mice to 8 weeks of chronic glucocorticoid treatment. After long-term exposure to dexamethasone, endogenous production of corticosterone was completely shut down in wild-type mice as expected, and was incompletely suppressed in Cry-deficient mice (Fig. 4e), suggesting that cryptochromes directly participate in glucocorticoid-receptor-mediated repression of glucocorticoid synthesis through the HPA axis. Furthermore, circulating adrenocorticotrophic hormone (ACTH) seems to be elevated in DKO mice (Supplementary Fig. 7), suggesting that cryptochromes may regulate steroidogenesis upstream of pituitary secretion of ACTH. Chronic glucocorticoid treatment caused significant fasting hyperglycaemia both in wild-type and DKO mice, but it was more severe in mice lacking cryptochromes (Fig. 4f). In addition, Cry-deficient mice exposed to chronic glucocorticoid treatment exhibited greater glucose intolerance than Cry-deficient mice treated with saline for the same time period, whereas wild-type mice had the same ability to respond to a mild glucose challenge whether they were exposed to saline or dexamethasone over the 8 weeks of treatment (Fig. 4g). Together, these data suggest that cryptochrome repression of glucocorticoid-receptor-mediated transcription is an important mechanism by which circadian clocks modulate metabolic physiology *in vivo*.

It has become increasingly clear that circadian clocks play an important role in optimizing the temporal coordination of tissue

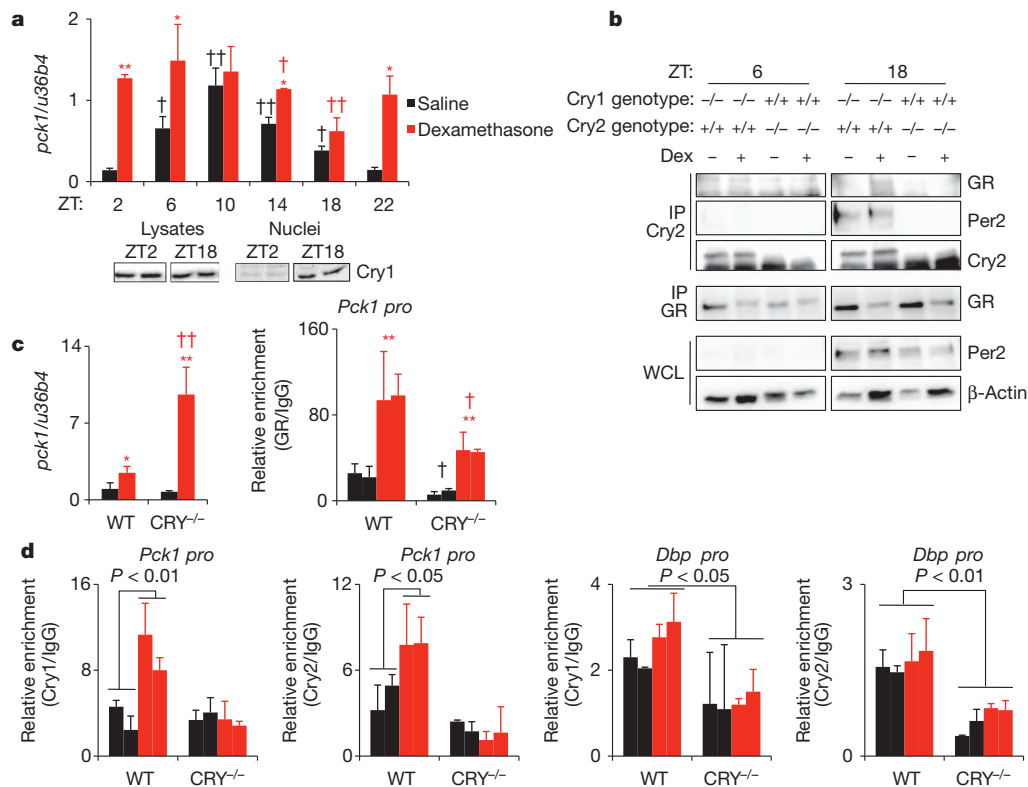


Figure 3 | Cryptochromes interact with glucocorticoid receptor on chromatin to regulate *pck1*. **a**, Top, *pck1* in mouse livers. * $P < 0.05$, ** $P < 0.01$, compared with saline; † $P < 0.05$, †† $P < 0.01$, compared with ZT2. Bottom, Cry1 immunoblots. **b**, Immunoblots of liver lysates or immunoprecipitations. **c**, Left, *pck1* at ZT16. Right, recovery of *Pck1* GRE.

* $P < 0.05$, ** $P < 0.01$, compared with saline; † $P < 0.05$, †† $P < 0.01$, compared with wild type (WT). **d**, Recovery of *Pck1* GRE or *Dbp* promoter E-box at ZT16. Data represent the mean \pm s.e.m. of three samples (a, c) or duplicate samples (d) analysed in triplicate.

physiology with metabolic demands that fluctuate over the course of the day². Although oscillating transcriptional activation of the target genes of *Clock* and *Bmal1* contributes to these physiological rhythms,

some transcriptional targets, including *pck1*, undergo circadian oscillations independently of *Clock* and *Bmal1* (ref. 10). Some of these transcripts are likely to be regulated by rhythms of fasting and feeding

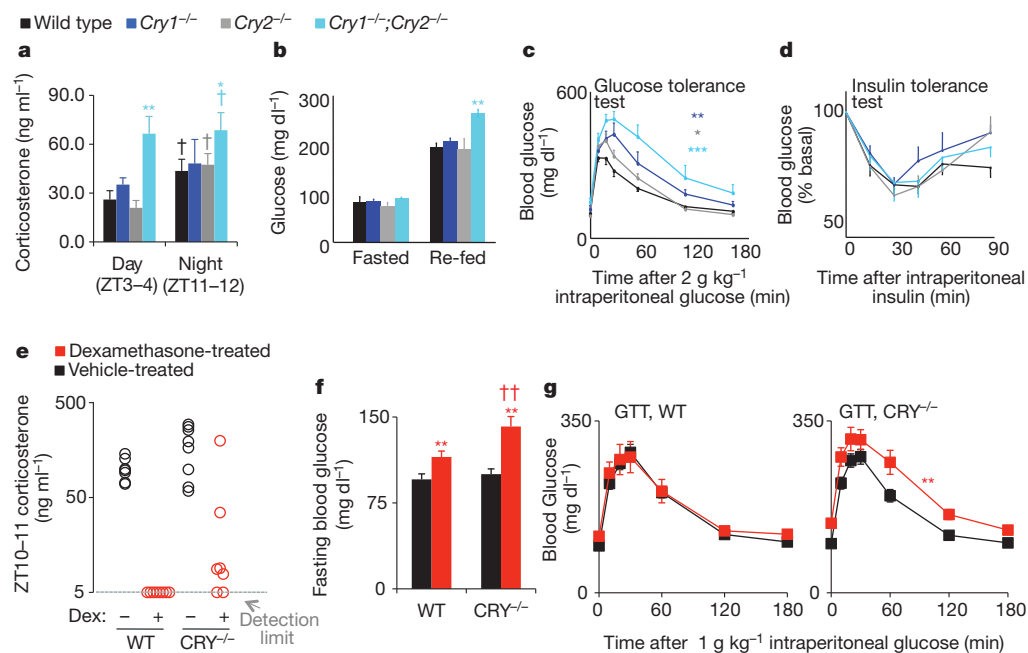


Figure 4 | Genetic loss of cryptochromes alters physiology. **a**, Serum corticosterone. † $P < 0.05$ compared with ZT3–4; * $P < 0.05$, ** $P < 0.01$, compared with wild type. **b**, Fasted and re-fed blood glucose. ** $P < 0.01$ compared with wild type. **c**, Glucose tolerance test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared with wild type. **d**, Insulin tolerance test.

e, Corticosterone in sera collected at ZT10–11. **f**, Fasted blood glucose. ** $P < 0.01$ compared with saline-treated animals; †† $P < 0.01$ compared with wild type. **g**, Glucose tolerance tests. ** $P < 0.01$ compared with saline-treated animals. In a–d, f, g, data represent the mean \pm s.e.m. for six to eight animals per group.

and subsequent oscillations in insulin, glucagon and other hormones²⁰; they may also be modulated by circadian repressors in the period and cryptochrome families.

In this study, we describe an interaction between cryptochromes and the glucocorticoid receptor, through which Cry1 and Cry2 repress glucocorticoid-receptor-dependent transcription, including the induction of *pck1* in the liver (Supplementary Fig. 1). We demonstrate that Cry1 and Cry2 interact with glucocorticoid receptor in mouse liver and are associated with a GRE in the *Pck1* promoter in a glucocorticoid-dependent manner. The physiological relevance of these associations is apparent in the increased susceptibility of cryptochrome-deficient mice to glucocorticoid-induced hyperglycaemia. Notably, *cry1*^{-/-}, *cry2*^{-/-} and *cry1*^{-/-}; *cry2*^{-/-} mice were also strikingly intolerant of a high dose of glucose, even in the absence of supplemental glucocorticoid treatment. This reflects a combination of increased glucocorticoid production, increased glucocorticoid-induced *pck1* transcription and other cryptochrome-dependent pathways, including both Clock/Bmal1-dependent transcriptional programs and recently reported Cry-dependent effects on cyclic AMP (cAMP)-responsive element-binding protein (CREB)-mediated transcription²¹. Past studies have indicated a role for PER2 upregulation in the hyperglycaemia induced by glucocorticoids through glucocorticoid receptor binding a GRE in the *Per2* promoter²²; we have found a distinct regulatory axis, involving direct physical association between cryptochromes and glucocorticoid receptor, by which circadian clocks modulate the response to glucocorticoids.

In addition, we found that Cry1 and Cry2 participate in glucocorticoid-dependent suppression of the HPA axis and the production of endogenous glucocorticoids. The observed effects on corticosterone production are reminiscent of a recent study in which *cry1*^{-/-}; *cry2*^{-/-} mice were found to have constitutively elevated production of aldosterone owing to adrenal de-repression of type VI 3-β hydroxysteroid dehydrogenase (*hsd3b6*)²³. Combined with our observation of direct interactions between cryptochromes and multiple steroid hormone receptors (Fig. 1a and data not shown), these data suggest that cryptochromes may play a general role in feedback inhibition of steroid hormone biosynthesis. However, the anatomic location of action may differ as previous studies did not observe a role of cryptochromes in corticosteroid biosynthetic gene expression in adrenal cortex, although they do seem to modulate aldosterone-synthesizing pathways in the adrenal gland²³.

These results suggest a specific mechanism through which cryptochromes couple the activity of clock and receptor target genes to complex genomic circuits underpinning normal metabolic homeostasis. Although we found that cryptochromes participate in glucocorticoid regulation of gluconeogenesis and steroidogenesis, they are not required for trans-repression of inflammatory gene transcription. Thus, the undesirable metabolic side effects of glucocorticoids used to suppress inflammation may be alleviated by altering the timing of treatment or by combining them with agents that can stabilize Cry1 and/or Cry2 in the liver. As metabolic syndrome is a complex blend of chronic inflammation and poorly controlled glucose homeostasis, the ability of cryptochromes to integrate metabolism and circadian rhythms suggests that targeting the glucocorticoid receptor/Cry interface could be a new therapeutic strategy.

METHODS SUMMARY

The Methods section provides detailed information about all experimental procedures including (1) the generation, culture and transfection of cell lines, (2) generation or sources of plasmids and antibodies, (3) luciferase assays, (4) preparation of samples for western blotting, immunoprecipitation and chromatin immunoprecipitation, (5) performance and analysis of microarray experiments, (6) quantitative PCR, (7) analysis of behaviour and metabolism in mice, (8) measurement of ACTH and corticosterone by immunoassay and (9) acute and chronic treatment of mice with synthetic glucocorticoids.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 March; accepted 4 November 2011.

Published online 14 December 2011.

- Yang, X. *et al.* Nuclear receptor expression links the circadian clock to metabolism. *Cell* **126**, 801–810 (2006).
- Green, C. B., Takahashi, J. S. & Bass, J. The meter of metabolism. *Cell* **134**, 728–742 (2008).
- Weinberger, C. *et al.* Identification of human glucocorticoid receptor complementary DNA clones by epitope selection. *Science* **228**, 740–742 (1985).
- Lonard, D. M. & O'Malley, B. W. The expanding cosmos of nuclear receptor coactivators. *Cell* **125**, 411–414 (2006).
- De Bosscher, K., Vanden Berghe, W. & Haegeman, G. Cross-talk between nuclear receptors and nuclear factor kappaB. *Oncogene* **25**, 6868–6886 (2006).
- Damiola, F. *et al.* Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14**, 2950–2961 (2000).
- Lamia, K. A. *et al.* AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* **326**, 437–440 (2009).
- Asher, G. *et al.* SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell* **134**, 317–328 (2008).
- Asher, G. *et al.* Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. *Cell* **142**, 943–953 (2010).
- Lamia, K. A., Storch, K. F. & Weitz, C. J. Physiological significance of a peripheral tissue circadian clock. *Proc. Natl Acad. Sci. USA* **105**, 15172–15177 (2008).
- Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**, 627–631 (2010).
- Sadacca, L. A., Lamia, K. A., deLemos, A. S., Blum, B. & Weitz, C. J. An intrinsic circadian clock of the pancreas is required for normal insulin release and glucose homeostasis in mice. *Diabetologia* **54**, 120–124 (2010).
- Storch, K. F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
- Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
- Kornmann, B., Schaad, O., Bujard, H., Takahashi, J. S. & Schibler, U. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol.* **5**, e34 (2007).
- Schmutz, I., Ripberger, J. A., Baeriswyl-Aebischer, S. & Albrecht, U. The mammalian clock component PERIOD2 coordinates circadian output by interaction with nuclear receptors. *Genes Dev.* **24**, 345–357 (2010).
- Sato, T. K. *et al.* A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* **43**, 527–537 (2004).
- Balsalobre, A. *et al.* Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* **289**, 2344–2347 (2000).
- Reichardt, H. M. *et al.* DNA binding of the glucocorticoid receptor is not essential for survival. *Cell* **93**, 531–541 (1998).
- Vollmers, C. *et al.* Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc. Natl Acad. Sci. USA* **106**, 21453–21458 (2009).
- Zhang, E. E. *et al.* Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. *Nature Med.* **16**, 1152–1156 (2010).
- So, A. Y.-L., Bernal, T. U., Pillsbury, M. L., Yamamoto, K. R. & Feldman, B. J. Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. *Proc. Natl Acad. Sci. USA* **106**, 17582–17587 (2009).
- Doi, M. *et al.* Salt-sensitive hypertension in circadian clock-deficient Cry-null mice involves dysregulated adrenal Hsd3b6. *Nature Med.* **16**, 67–74 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Kaufman for assistance with glucose tolerance tests, J. Alvarez for RNA sample preparation, H. Juguilon for luciferase assays, J. Vaughan for corticosterone and ACTH measurements, antigenic peptide design and peptide-KLH coupling reactions, H. Cho for sharing unpublished plasmids and R. Shaw for comments on the manuscript. This work was supported by National Institutes of Health grants DK057978 and DK062434 (to R.M.E.) and DK090188 (to K.A.L.), by support from the Glenn Foundation for Aging Research (to R.M.E. and K.A.L.), from the Helmsley Trust (to R.M.E.) and by a Merck fellowship from the Life Sciences Research Foundation (to K.A.L.).

Author Contributions K.A.L. and R.E. conceived the project and designed the research. K.A.L., S.J.P., G.D.B. and N.H.U. performed the experiments. J.W.J. provided critical reagents. K.A.L. and R.T.Y. analysed the data. K.A.L. and R.M.E. wrote the paper. All authors edited the manuscript.

Author Information Microarray data are deposited in Gene Expression Omnibus under accession number GSE24469. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.M.E. (evans@salk.edu) or K.A.L. (klamia@scripps.edu).

METHODS

Cells and cell culture. HEK 293T cells were purchased from the American Type Culture Collection and grown in complete Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen) supplemented with 10% dialysed fetal bovine serum, penicillin and streptomycin in an incubator at 37 °C and maintained at 5% CO₂. MEFs stably expressing Flag-Cry1 were as previously described⁶. Primary wild-type and *cry1*^{-/-}; *cry2*^{-/-} MEFs were generated from embryonic day 15.5 embryos derived from mating *cry1*^{+/-}; *cry2*^{+/-} male mice with *cry1*^{+/-}; *cry2*^{+/-} female mice. A pregnant female was euthanized by CO₂ inhalation, her uteri were dissected and rinsed in PBS, and each embryo was extracted from the surrounding tissue. After removal of the heads and fetal livers, the remaining embryonic tissue was minced with razor blades, digested with trypsin and re-suspended in DMEM containing 15% FBS to generate a homogeneous cell slurry, which was plated in 6 cm dishes. MEFs were expanded slowly and were used for experiments after 8–10 passages. Genomic DNA isolated from the fetal livers was used to genotype the resulting cell lines. Primary macrophages were isolated and cultured as described²⁴.

Plasmids and transfection. pcDNA3-2×Flag-mCRY1, pcDNA3-2×Flag-mCRY2, pcDNA3-Myc-Cry1 and pcDNA3-mPER2 were gifts from C. Weitz; pcDNA3-Flag-Rev-Erb-a was a gift from H. Cho; pcDNA3.1-RORα-v5, pcDNA3.1-RORβ-v5, pcDNA3.1-RORγ-v5, pcDNA3-GR-v5, pcDNA3-AR-v5, pcDNA3-ERR1-v5, pcDNA3-ERR2-v5 and pcDNA3-ERR3-v5 were gifts from J. Jonker; hGRα, hGRαΔ262–404, hGRαΔ490–515 and hGRαΔ532–697 were as previously described³. Transfections were performed using FuGene HD (Roche).

Luciferase assays. CV-1 cells were transfected with the indicated plasmids and treated with either vehicle (ethanol) or 1 μM dexamethasone 24 h after transfection. After overnight treatment, cells were lysed in passive lysis buffer (Promega) and analysed using Dual-Glo Luciferase Reporter System (Promega). Firefly luciferase signal was normalized to β-galactosidase. All luciferase assay data represent the mean ± s.e.m. of triplicate samples.

Preparation of protein extracts, immunoprecipitation, and immunoblotting. Cell and liver extracts were prepared in lysis buffer containing 1% Triton X-100 (cell extracts) or 1% NP-40 (liver extracts) as previously described²⁵. Immunoprecipitation antibodies were M2-agarose (Sigma A2220), PER21A (Alpha Diagnostic International), anti-GR M-20 (Santa Cruz Biotechnology sc-1004) and polyclonal antibodies raised in guinea pigs against the C termini of Cry1 (amino acids 583–606) or Cry2 (amino acids 563–592) crosslinked to KLH and affinity purified against the immunogenic peptides crosslinked to a column (AminoLink Immobilization Kit, Thermo Scientific 44890). Antibodies used for western blotting were anti-v5, anti-myc and anti-Flag polyclonal antibodies (Sigma), PER21A, GR135 antiserum³ and anti-Cry1-CT and anti-Cry2-CT as described above.

Microarrays. Total RNA was extracted using Trizol reagent (Invitrogen); purity of the RNA was assessed by Agilent 2100 Bioanalyser. Five hundred nanograms of RNA were reverse transcribed into antisense RNA and biotin-UTP labelled using an Illumina TotalPrep RNA Amplification Kit (Ambion). Antisense RNA was quantified using an Agilent Bioanalyser 2100 and hybridized to the Illumina mouseRefseq-8v2 Expression BeadChip using standard protocols (Illumina). Image data were converted into unnormalized Sample Probe Profiles using the Illumina BeadStudio software and analysed on the VAMPIRE microarray analysis framework. Stable variance models were constructed for each of the experimental conditions ($n = 2$). Differentially expressed probes were identified using the unpaired VAMPIRE significance test with a two-sided, Bonferroni-corrected threshold of $\alpha_{\text{Bonf}} = 0.05$. The VAMPIRE statistical test is a Bayesian statistical method that computes a model-based estimate of noise at each level of gene expression. This estimate was then used to assess the significance of apparent differences in gene expression between two experimental conditions. Lists of altered genes generated by VAMPIRE were mapped to pathways using the VAMPIRE tool GOby to determine whether any KEGG categories were overrepresented using a Bonferroni error threshold of $\alpha_{\text{Bonf}} = 0.05$. A heat map was constructed with cubic spline-normalized values using the CIMminer program at <http://discover.nci.nih.gov/>, a development of the Genomics and Bioinformatics Group, Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute.

Gene expression. RNA was extracted from livers or cultured cells with Trizol or by using the Qiagen RNeasy purification system. cDNA was prepared using the SuperscriptII reverse transcriptase (Invitrogen) and analysed for gene expression using quantitative real-time PCR with SYBR green (Invitrogen SybrGreenER or Biorad iQ SybrGreen supermix) chemistry. Primer sequences are available upon

request. All quantitative PCR data represent the mean ± s.e.m. for three samples of each condition, analysed in triplicate.

Chromatin immunoprecipitation. Freshly dissected mouse livers were rinsed in ice-cold PBS, homogenized in 167 ng ml⁻¹ DSG in PBS using a Dounce homogenizer, and incubated at room temperature for 30 min. The homogenates were centrifuged at 3000g for 5 min and the pellets were re-suspended in 1% formaldehyde in PBS and incubated at room temperature for 10 min. Crosslinking was stopped by addition of glycine to a final concentration of 100 mM and samples were placed on ice. After cross-linking, nuclei were purified using a sucrose gradient by the method of Lavery and Schibler²⁶. Purified nuclei were washed three times in ice-cold PBS and re-suspended in buffer containing 150 mM NaCl, 5 mM EDTA, 50 mM Tris pH 7.5, 0.5% NP-40 and 1% Triton X-100; chromatin immunoprecipitation was continued as described by Nelson *et al.*²⁷ using a modified shearing buffer (1% SDS, 10 mM EDTA, 50 mM Tris) and 3 μg of immunoprecipitating antibody per reaction. Primers used for the detection of the Dbp promoter E-box region were as described²⁸. All chromatin immunoprecipitation data represent the mean ± s.e.m. for duplicate samples each analysed in triplicate.

Mice. *Cry1*^{-/-}; *Cry2*^{-/-} mice were from A. Sancar²⁹. Indirect calorimetry studies were conducted in a Comprehensive Lab Animal Monitoring System (eight-chamber system, Columbus Instruments). All animal care and treatments were in accordance with the Salk Institute guidelines for the care and use of animals.

Corticosterone and ACTH measurements. Corticosterone and ACTH were measured in mouse serum collected by tail bleed using the Corticosterone Double Antibody 125I RIA Kit (MP Biomedicals 7120103) or ACTH 125I RIA Kit (Diasorin catalogue 24130).

Fasted and re-fed blood glucose measurements. Mice were placed in clean cages (without food) and fasted from ZT11 to ZT2. For re-feeding, food was returned to the cages for 2 h from ZT2 to ZT4. Glucose was measured in blood collected from the tail vein using a OneTouch Basic glucometer.

Glucose tolerance test. Mice were placed in clean cages (without food) at ZT10–11 1 day before the experiment and were injected with glucose (1 or 2 mg g⁻¹ body-weight; 20% glucose in 0.9% NaCl). Blood glucose levels were measured before the injection of glucose and at 15, 30, 60, 120 and 180 min after injection.

Insulin tolerance test. Mice were placed in clean cages (without food) at ZT4 and were injected intraperitoneally 2 h later with 0.5–1.0 units per kilogram body-weight of Novolin-R in 0.9% NaCl. Blood glucose was measured using a One Touch Basic glucometer (Lifescan) before injection of insulin and at 15, 30, 45, 60 and 90 min after insulin injection.

Acute dexamethasone treatment. For experiments measuring the *in vivo* transcriptional response to dexamethasone (Fig. 3), mice were treated with 1 mg dexamethasone per kilogram bodyweight by intraperitoneal injection 1 h before dissection. All data represent the mean ± s.e.m. for three animals treated with saline (–) or dexamethasone (+), with each sample analysed in triplicate.

Chronic dexamethasone treatment. Mice were treated with 1 mg dexamethasone per gram bodyweight by intraperitoneal injection of an equimolar amount of dexamethasone 21-phosphate sodium salt (Sigma D1159) dissolved in sterile saline every other day at ZT11–12 for 8 weeks. Mice in the control group were injected with sterile saline at the same time.

Statistics. Luciferase assay data were analysed using a two-tailed *t*-test. Quantitative PCR data were analysed by analysis of variance (Fig. 3a) or a two-tailed *t*-test. Glucose and insulin tolerance test data were analysed using repeated-measures analysis of variance for effects of genotype × time or treatment × time. Corticosterone and glucose measurements were analysed by a two-tailed *t*-test.

24. Barish, G. D. *et al.* A nuclear receptor atlas: macrophage activation. *Mol. Endocrinol.* **19**, 2466–2477 (2005).
25. Lamia, K. A. *et al.* Increased insulin sensitivity and reduced adiposity in phosphatidylinositol 5-phosphate 4-kinase β mice. *Mol. Cell. Biol.* **24**, 5080–5087 (2004).
26. Lavery, D. J. & Schibler, U. Circadian transcription of the cholesterol 7 α hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.* **7**, 1871–1884 (1993).
27. Nelson, J. D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature Protocols* **1**, 179–185 (2006).
28. Stratmann, M., Stadler, F., Tamanini, F., van der Horst, G. T. & Ripperger, J. A. Flexible phase adjustment of circadian albumin D site-binding protein (DBP) gene expression by CRYPTOCHROME1. *Genes Dev.* **24**, 1317–1328 (2010).
29. Thresher, R. J. *et al.* Role of mouse cryptochrome blue-light photoreceptor in circadian photoreponses. *Science* **282**, 1490–1494 (1998).

GlcNAcylation of histone H2B facilitates its monoubiquitination

Ryoji Fujiki¹, Waka Hashiba¹, Hiroki Sekine¹, Atsushi Yokoyama¹, Toshihiro Chikanishi¹, Saya Ito¹, Yuuki Imai¹, Jaehoon Kim², Housheng Hansen He³, Katsuhide Igarashi⁴, Jun Kanno⁴, Fumiaki Ohtake¹, Hirochika Kitagawa¹, Robert G. Roeder², Myles Brown³ & Shigeaki Kato^{1,5}

Chromatin reorganization is governed by multiple post-translational modifications of chromosomal proteins and DNA^{1,2}. These histone modifications are reversible, dynamic events that can regulate DNA-driven cellular processes^{3,4}. However, the molecular mechanisms that coordinate histone modification patterns remain largely unknown. In metazoans, reversible protein modification by *O*-linked *N*-acetylglucosamine (GlcNAc) is catalysed by two enzymes, *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA)^{5,6}. However, the significance of GlcNAcylation in chromatin reorganization remains elusive. Here we report that histone H2B is GlcNAcylated at residue S112 by OGT *in vitro* and in living cells. Histone GlcNAcylation fluctuates in response to extracellular glucose through the hexosamine biosynthesis pathway (HBP)^{5,6}. H2B S112 GlcNAcylation promotes K120 monoubiquitination, in which the GlcNAc moiety can serve as an anchor for a histone H2B ubiquitin ligase. H2B S112 GlcNAc was localized to euchromatic areas on fly polytene chromosomes. In a genome-wide analysis, H2B S112 GlcNAcylation sites were observed widely distributed over chromosomes including transcribed gene loci, with some sites co-localizing with H2B K120 monoubiquitination. These findings suggest that H2B S112 GlcNAcylation is a histone modification that facilitates H2BK120 monoubiquitination, presumably for transcriptional activation.

Some nuclear proteins have been shown to be GlcNAcylated by OGT, for example the enzymatic activity of histone H3K4 methyltransferase 5 (MLL5) is modulated by GlcNAcylation^{7–9}. To identify chromatin substrates for OGT further, we screened for unknown GlcNAcylated glycoproteins in HeLa cell chromatin. GlcNAcylated proteins were purified by WGA lectin column chromatography and anti-GlcNAc antibody (clone RL2). Liquid chromatography–mass spectrometry (LC–MS)/MS analysis of the fraction revealed 284 factors, including previously reported GlcNAcylated glycoproteins^{5,10} (Supplementary Table 1). Among the candidates, the enrichment of nucleosomes was confirmed by silver staining and western blotting (Supplementary Fig. 2), suggesting one or more histone(s) might have been GlcNAcylated. As OGT is the only known nuclear enzyme for protein GlcNAcylation⁵, we asked whether histones served as substrates for OGT *in vitro* (Supplementary Fig. 3). H2A and H2B, as well as H2A variants (H2A.X and H2A.Z), but not H3 and H4, appeared to be GlcNAcylated (Fig. 1a). With histone octamers, H2B, but not H2A, appeared to serve as a substrate (Fig. 1b). Likewise, H2B in *Drosophila* histone was also GlcNAcylated (Supplementary Fig. 4), implying that H2B GlcNAcylation is conserved in metazoans.

A quadrupole (Q)-time of flight (TOF) MS assessment of the *in vitro* GlcNAcylated H2B showed that OGT could transfer three GlcNAc moieties to H2B (Supplementary Fig. 5). Electro-transfer-dissociation (ETD)–MS/MS mapped the sites to S91, S112 and S123 (Fig. 1c and Supplementary Fig. 6). Unlike a recent report¹¹, we were unable to

detect the reported sites in H2B S36 and H4 S47. However, H2A T101 was detected as a GlcNAc site when H2A protein alone was used (data not shown). This discrepancy in identified GlcNAc sites might be due to differences in experimental approaches.

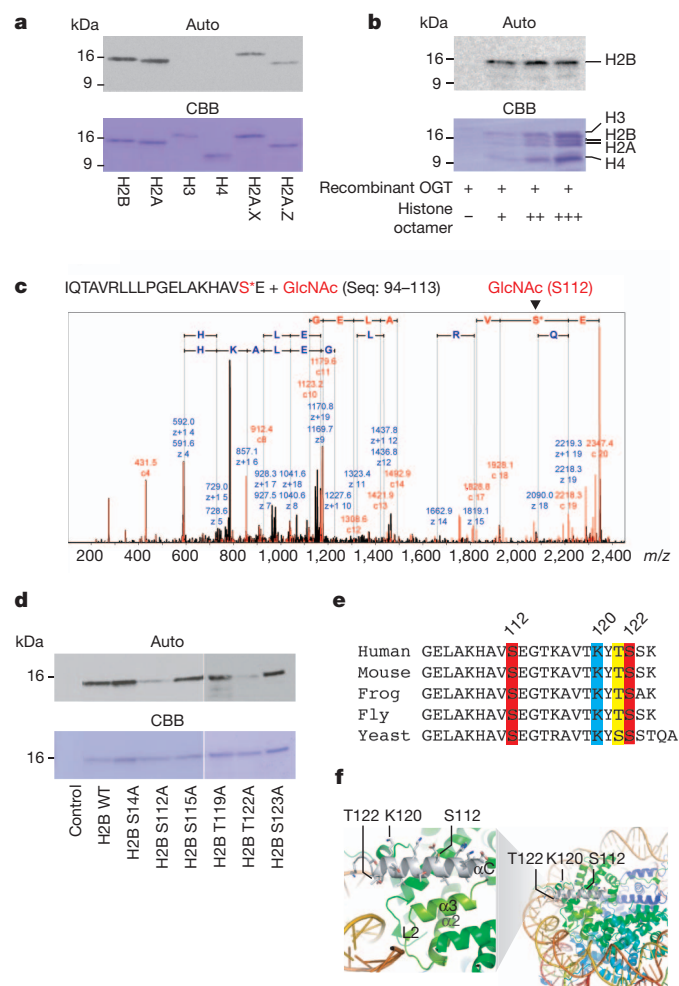


Figure 1 | H2B is GlcNAcylated at the C-terminal S112. **a**, **b**, *In vitro* OGT assay with recombinant histones (**a**) or the octamers reconstituted *in vitro* (**b**). Histones were GlcNAcylated by uridine diphosphate (UDP)-[³H]GlcNAc and OGT, and the radiolabelled histones were subjected to autoradiography (top) and CBB staining (bottom). **c**, ETD–MS/MS scanned the GlcNAcylated peptides (2349.43 *m/z*) in Supplementary Fig. 5b. **d**, A series of H2B mutants at the indicated S/T was assessed by *in vitro* OGT assays. **e**, Sequence alignment of α C. **f**, The locations of the GlcNAc sites and the ubiquitination site of H2B in a nucleosome. The α C helix is illustrated as a white ribbon.

¹Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. ²Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, New York 10065, USA. ³Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Division of Cellular and Molecular Toxicology, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. ⁵ERATO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan.

Next, *in vitro* OGT assays using peptide arrays covering full-length H2B revealed peaks at 101–115 peptides in the carboxy (C)-terminal α -helix (α C)¹² (Supplementary Fig. 7). This peptide was found to bear only one moiety by matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF)/MS (Supplementary Fig. 8). Indeed, substitutions of S112 and T122 to A significantly reduced *in vitro* GlcNAcylation by OGT (Fig. 1d), but not mutations in the amino (N)-terminal tail (Supplementary Fig. 9). On the basis of these data, we concluded that the conserved S112 was a GlcNAc site in H2B, whereas T122 might be needed for recognition by OGT (Fig. 1e, f).

With our newly developed antibody (Supplementary Fig. 10), H2B S112 GlcNAc was detected in histones of HeLa cells. Depletion of glucose from the media for 24 h induced deglycosylation with neither overt cell death (Fig. 2a and Supplementary Fig. 11) nor alteration in histone acetylation marks of cell state indicators (H3 K14, H3 K56, H4 K16)^{13,14} (Supplementary Fig. 12). H2B S112 GlcNAc could be restored by re-treatment with glucose at physiological concentrations (Supplementary Fig. 13).

Because many histone modifications are orchestrated, we tested if H2B S112 GlcNAc influenced H2B K120 monoubiquitination because

of their proximity. After glucose depletion, replenishment of glucose gradually increased global GlcNAcylation of proteins, followed by H2B S112 GlcNAc and H2B monoubiquitination (Fig. 2b, c). Their reciprocal modifications disappeared when OGT was knocked down (Fig. 2d and Supplementary Fig. 14). In addition, in the immunoprecipitates of H2B containing the S112A and T122A double mutations (H2B AA), no response of K120 monoubiquitination to extracellular glucose was detected (Fig. 2e and Supplementary Fig. 15). Conversely, GlcNAcylation of H2B S112 was observed, even when K120 was mutated to R (Fig. 2e). From these findings, we conclude that H2B K120 monoubiquitination is mediated, at least in part, through S112 GlcNAcylation.

As glucosamine, but not pyruvate, potentiated H2B S112 GlcNAc (Fig. 2f), it appeared that this GlcNAcylation step was dependent on the HBP. To clarify this point, two HBP inhibitors (DON and AZA) were tested (Supplementary Information). After glucose depletion from media, these inhibitors attenuated the effect of glucose in H2B S112 GlcNAcylation along with K120 monoubiquitination (Fig. 2f).

In yeast, it was previously shown that H2B K120 monoubiquitination was induced by carbohydrates by glycolysis¹⁵. To address this issue, inhibitors of both glycolysis and deGlcNAcylation were applied to assess the crosstalk between the two modifications. When the cells were treated with iodoacetate, which blocks glycolysis but not HBP¹⁵, the glucose effects on histone modifications were impaired, whereas the additional treatment of an OGA inhibitor (PUGNac) restored both H2B S112 and K120 monoubiquitination (Supplementary Fig. 16). These data support the notion that H2B S112 GlcNAc senses decreases in glucose levels below normal levels and acts to promote H2B monoubiquitination, a modification that is associated with active transcription. Together with the fact that OGT is absent in yeast⁶, the present H2B S112 GlcNAc-dependent pathway appears to constitute a system capable of sensing nutritional states in metazoans.

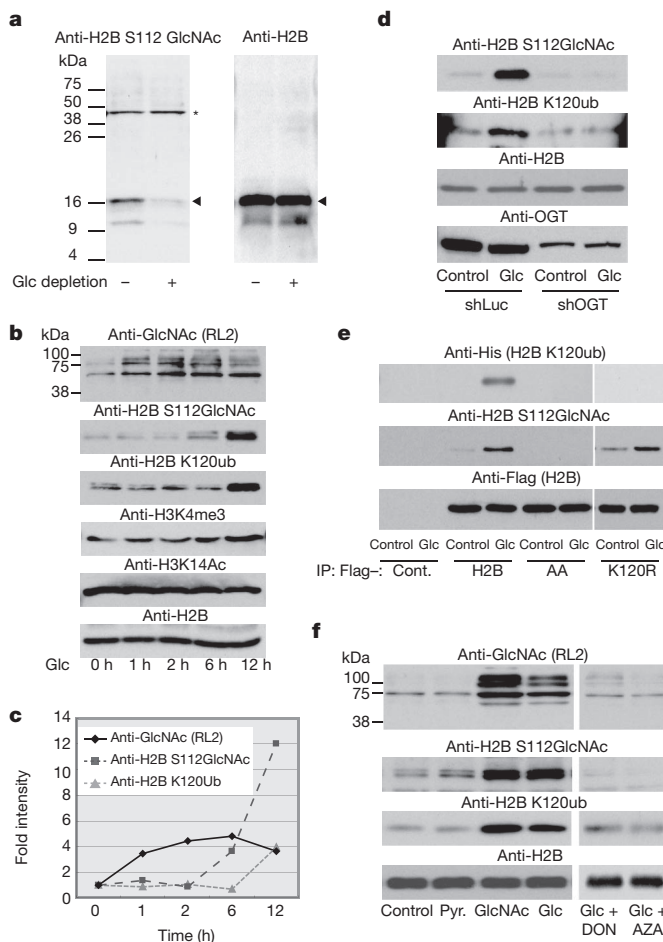


Figure 2 | H2B S112 GlcNAc is a glucose-responsive modification linked to K120 monoubiquitination (ub). **a**, Chromatin was prepared from HeLa cells cultured in media with or without 1 g l⁻¹ glucose (Glc) for 24 h, and subjected to western blotting. Arrowheads show the indicated proteins. Asterisks indicate non-specific band. **b**, **c**, After 24 h Glc depletion, chromatin samples were prepared from HeLa cells treated with 4.5 g l⁻¹ Glc for the indicated time. The intensities of the western blotting bands (**b**) were quantified (**c**). **d**, **e**, The effects of OGT knockdown (**d**) or H2B mutations (**e**) on H2B modifications after Glc replenishment. **f**, Western blotting analysis of the H2B modifications in HeLa cells that were cultured in DMEM without Glc (Cont.), or supplemented with 1 mM pyruvate (Pyr.), 10 mM GlcNAc or 4.5 g l⁻¹ Glc with or without HBP inhibitors, 6-diazo-5-oxo-L-norleucine (100 μ M, DON) or azaserine (100 μ M, AZA).

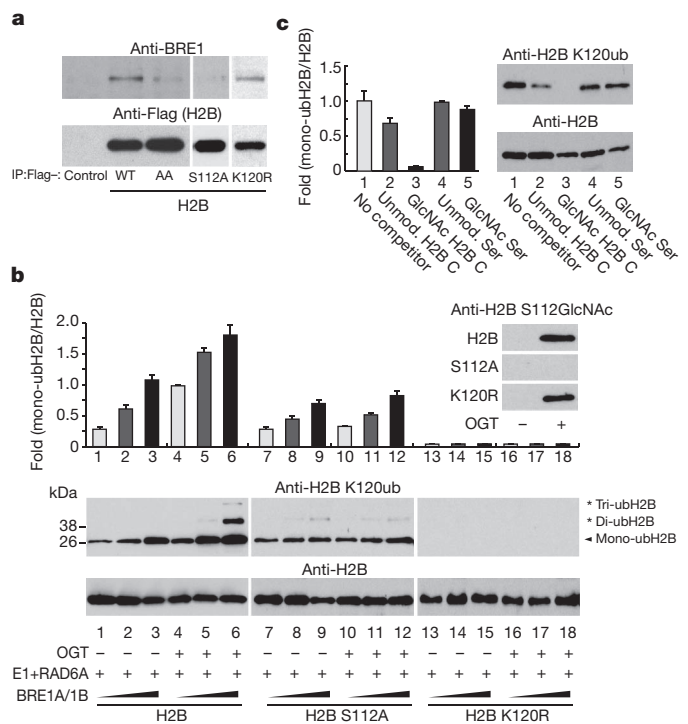


Figure 3 | GlcNAcylation at S112 facilitates ubiquitination at K120 in H2B. **a**, Western blotting analysis of the interaction of H2B mutants with BRE1A. **b**, **c**, *In vitro* monoubiquitination assay with GlcNAcylation H2B (**b**), or in the presence of competitor peptides (**c**). H2B was GlcNAcylation *in vitro* (**b**, top right), and the reactants were subsequently ubiquitinated by H2B monoubiquitination ligase. The reaction was performed with the indicated competitor peptides (0.25 μ g ml⁻¹) (**c**). H2B K120 monoubiquitination was detected by western blotting (**b**, bottom; **c**, right) and quantified (**b**, top; **c**, left). Error bars, means and s.d. ($n = 3$).

The terminal GlcNAc of polysaccharides reportedly serves as a recognition moiety for E3 monoubiquitination ligase¹⁶. Therefore, we proposed that H2B S112 GlcNAc affected K120 monoubiquitination by the BRE1A/1B complex¹⁷. Flag-tagged H2B, but not AA or S112A, was co-immunoprecipitated with BRE1A (Fig. 3a). This association was observed in the presence of physiological levels of glucose in the media, and BRE1A was bound to H2B S112 GlcNAc (Supplementary Fig. 17). We then assessed how the GlcNAcylation of H2B influenced its *in vitro* ubiquitination by E1, RAD6A (E2) and the BRE1A/1B complex (E3). Although H2B K120 could be substantially ubiquitinated only by the ligases (Supplementary Fig. 18), GlcNAcylation of H2B promoted subsequent H2B ubiquitination, but not its S112A mutant (Fig. 3b). Likewise, ubiquitination was significantly attenuated by the presence of an H2B-S112-GlcNAcylated peptide, but not by either the unmodified control peptide or by GlcNAcylated serine (Fig. 3c). On the basis of these results, we conclude that the GlcNAc moiety at H2B S112 may anchor H2B monoubiquitination ligase.

To illustrate the role of H2B S112 GlcNAc in chromatin regulation, its location was visualized on fly polytene chromosomes. H2B S112 GlcNAc was detected widely in euchromatin, and, as anticipated, its signal disappeared in an OGT-disrupted fly, *sxc¹/sxc⁷⁸* (Supplementary Fig. 19). H2B S112 GlcNAc overlapped with H3K4 me2 more than with H3K9 me2 or H3K27 me3 (Fig. 4a). Similarly, in immunostained HeLa cells, H2B S112 GlcNAc sites appeared exclusively in 4',6-diamidino-2-phenylindole (DAPI)-poor areas (Supplementary Fig. 20).

Thus, H2B S112 GlcNAc probably accumulates in active chromatin rather than inactive chromatin.

To determine the precise loci of H2B S112 GlcNAc in HeLa cells, we performed chromatin immunoprecipitation (ChIP) and high-throughput sequencing (ChIP-seq). We confirmed ChIP quality by enrichments of H2B GlcNAc as well as H3K4 me2 and H2B K120 monoubiquitination, but neither H3K9 me2 nor H3K27 me3 (Supplementary Fig. 21). A total of 47,375 peaks were found widely distributed over the genome (Supplementary Fig. 22). However, H2B S112 GlcNAc peaked near transcription start sites (TSS), whereas the distribution decreased at transcription termination sites (TTS) (Fig. 4b), suggesting that it correlated with transcriptional regulation. To test this assumption, the activities of genes harbouring H2B S112 GlcNAc near TSS were estimated by microarray analysis (Supplementary Table 2). The average profiles near TSS significantly correlated with gene activity (Fig. 4c). Moreover, the expression levels of the 1,299 genes were reliably measured, and 1,021 genes showed high expression (Supplementary Fig. 23a and Supplementary Table 3b). Moreover, gene ontology analysis revealed that there was an association of the genes harbouring H2B S112 GlcNAc to cellular metabolic processes (Supplementary Fig. 23b and Supplementary Table 3c).

Next, we analysed the genome-wide overlap of H2B S112 GlcNAc with K120 monoubiquitination. A total of 44,158 peaks of H2B K120 monoubiquitination were detected, and their average profiles near TSS were similar to those profiles of H2B S112 GlcNAc (Supplementary

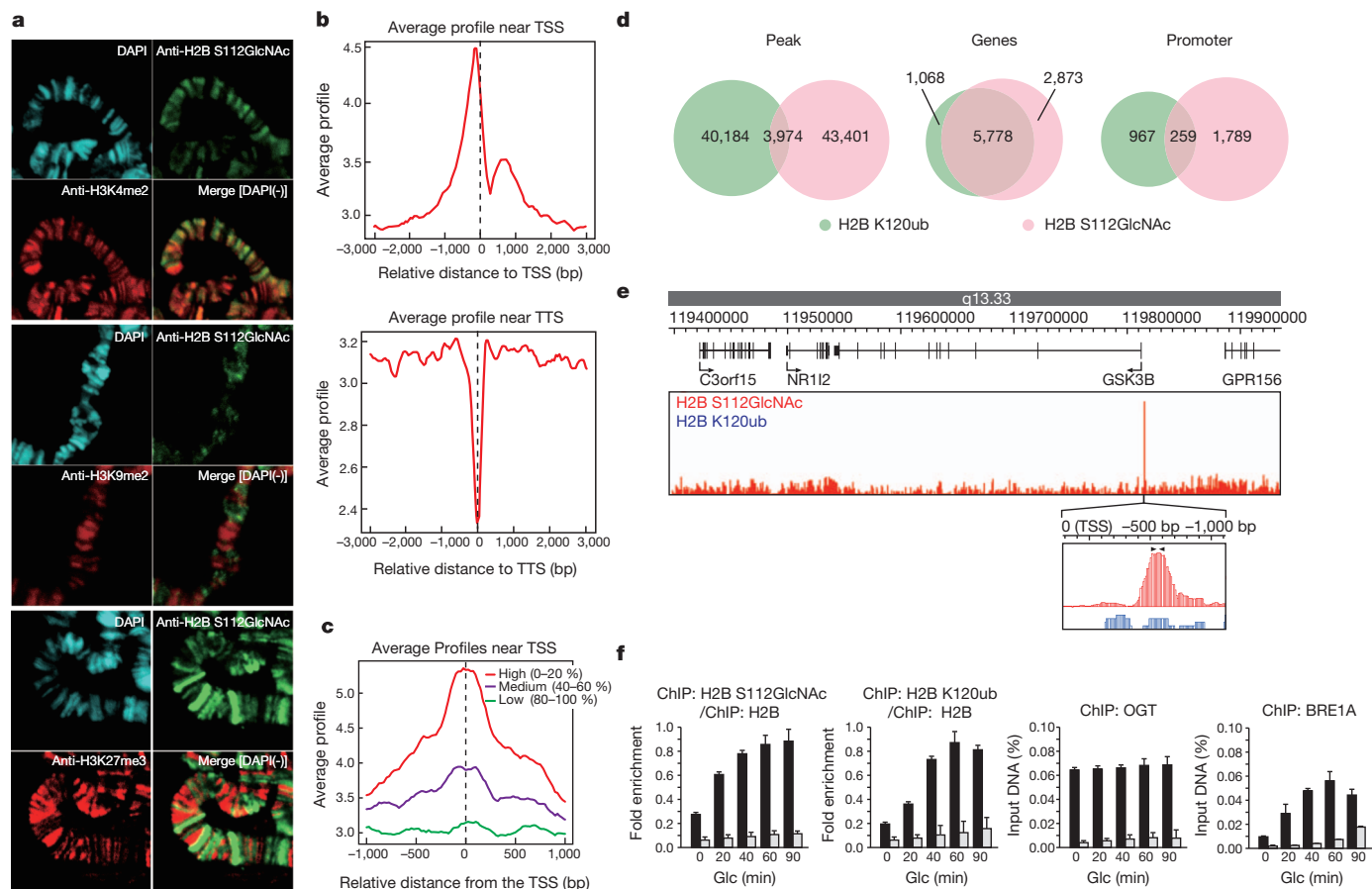


Figure 4 | GlcNAcylated H2B is associated with transcribed genes. **a**, Polytene staining with α -H2B S112 GlcNAc (green) and DAPI (blue) along with α -H3K4me2 (red, top), α -H3K9me2 (red, middle) or α -H3K27me3 (red, bottom). **b–e**, ChIP-seq analysis of the H2B S112 GlcNAc and K120 monoubiquitination. The distributions of H2B S112 GlcNAc were averaged near TSS (top) and TTS (bottom) (**b**). The average profiles of H2B S112 GlcNAc near TSS were calculated based on the associated gene activities (**c**). Venn diagrams

showing overlap of the peaks (**d**, left), and the genes (**d**, middle) and the promoter (**d**, right) harbouring the modifications. The ChIP-seq profile surrounding the *GSK3B* gene (**e**). Arrowhead, position of qPCR primer. **f**, ChIP-qPCR validation in the *GSK3B* promoter. After Glc depletion, the control HeLa cells (black bar) and the OGT-knockdown cells (white bar) were replenished with Glc for 24 h. Then, the cells were subjected to ChIP with the indicated antibody and qPCR analysis. Error bars, means and s.d. ($n = 3$).

Fig. 24). Among the H2B K120 monoubiquitination peaks, nearly 10% (3,974 peaks) overlapped with H2B S112 GlcNAc peaks (Fig. 4d, left), and this evaluation was confirmed by a sequential ChIP–reChIP assay (Supplementary Fig. 25). Although 5,778 genes (66.8% of H2B S112 GlcNAc and 84.4% of K120 monoubiquitination) were found at the same loci (Fig. 4d, middle, and Supplementary Table 3d), 259 genes were identified when the two peaks were compared only within the promoters (Fig. 4d, right). The results of the ChIP-seq analysis were validated by ChIP–quantitative PCR (qPCR) assessment for the glycogen synthase kinase 3 β (*GSK3B*) gene (Fig. 4e, f). These findings suggest that at several H2B S112 GlcNAc sites, it aids H2B monoubiquitination ligase recruitment whereas at others additional or different factors may be operational.

Here we provide evidence that histone GlcNAcylation is a post-translational modification correlated with active transcriptional events, and is responsive to serum glucose levels and/or cellular energy states in certain cell types (Supplementary Fig. 1). Using an antibody that specifically recognizes the S112 GlcNAc moiety of endogenous H2B, H2B was shown to serve as an OGT substrate. We have focused on the role of H2B S112 GlcNAcylation in gene regulation (Supplementary Fig. 1). Genome-wide analysis revealed that H2B S112 GlcNAc was frequently located near transcribed genes, suggesting that histone GlcNAcylation facilitates transcription of the genes. This idea is supported by previous reports that transcriptional output driven by several transcription factors is co-activated by OGT^{9,18–20}. However, recent papers reported that *Drosophila* OGT is itself a polycomb group protein^{8,21}, and that many O-GlcNAcylated factors are involved in transcriptional repression and gene silencing^{7,8}. In this respect, it will be interesting to identify other histone glycosylation sites and investigate their roles in transcriptional repression as well as activation.

METHODS SUMMARY

Plasmids and cell culture. All plasmids were generated with standard protocols (see Methods). Retrovirus production, infection and sorting of the infected cells followed previously reported protocols⁹.

Purification of GlcNAc proteins from chromatin. Chromatin pellets were prepared from HeLa cells as previously described²². GlcNAc proteins were enriched with α -O-GlcNAc (RL2) antibody (Abcam) immobilized on Dynabeads (Invitrogen), and released with GlcNAc-O-serine.

Generation of monoclonal antibody. The synthetic H2B S112 GlcNAc peptide (CKHAV S(GlcNAc) EGTK) was used to immunize mice. The hybridomas were selected by enzyme-linked immunosorbent assay (ELISA) and western blotting analysis.

In vitro OGT and monoubiquitination assays. Flag-OGT, Flag-E1, and Flag-BRE1A/BRE1B were purified by baculoviral systems, whereas histones and 6 \times His-RAD6A were prepared from bacteria as previously reported^{17,23}. H2B was incubated with OGT or H2B monoubiquitination ligases *in vitro*, and its modification was detected by western blotting as previously reported^{9,23}.

ChIP-seq and ChIP–qPCR. ChIP and ChIP-seq library construction was performed as previously described^{24,25}, and the libraries were sequenced to 50 base pairs (bp) with HiSeq2000 (Illumina). The fragments of interest in the libraries were quantified with specific promoter sets (Methods) by qPCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July 2010; accepted 20 October 2011.

Published online 27 November 2011.

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

2. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
3. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
4. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).
5. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked β -N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017–1022 (2007).
6. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the ‘O-GlcNAc code’. *Sci. STKE* **2005**, re13 (2005).
7. Yang, X., Zhang, F. & Kudlow, J. E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110**, 69–80 (2002).
8. Gambetta, M. C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase *sxc/Ogt* in polycomb repression. *Science* **325**, 93–96 (2009).
9. Fujiki, R. *et al.* GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature* **459**, 455–459 (2009).
10. Wang, Z. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3**, ra2 (2010).
11. Sakabe, K., Wang, Z. & Hart, G. W. β -N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc. Natl Acad. Sci. USA* **107**, 19915–19920 (2010).
12. Luger, K. *et al.* Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
13. Das, C., Lucia, M. S., Hansen, K. C. & Tyler, J. K. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459**, 113–117 (2009).
14. Dang, W. *et al.* Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802–807 (2009).
15. Dong, L. & Xu, C. W. Carbohydrates induce mono-ubiquitination of H2B in yeast. *J. Biol. Chem.* **279**, 1577–1580 (2004).
16. Yoshida, Y. *et al.* E3 ubiquitin ligase that recognizes sugar chains. *Nature* **418**, 438–442 (2002).
17. Kim, J. *et al.* RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459–471 (2009).
18. Dentin, R. *et al.* Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* **319**, 1402–1405 (2008).
19. Chikanishi, T. *et al.* Glucose-induced expression of MIP-1 genes requires O-GlcNAc transferase in monocytes. *Biochem. Biophys. Res. Commun.* **394**, 865–870 (2010).
20. Jackson, S. P. & Tjian, R. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**, 125–133 (1988).
21. Sinclair, D. A. *et al.* *Drosophila* O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (*sxc*). *Proc. Natl Acad. Sci. USA* **106**, 13427–13432 (2009).
22. Sawatsubashi, S. *et al.* A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. *Genes Dev.* **24**, 159–170 (2009).
23. Fujiki, R. *et al.* Ligand-induced transrepression by VDR through association of WSTF with acetylated histones. *EMBO J.* **24**, 3881–3894 (2005).
24. He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nature Genet.* **42**, 343–347 (2010).
25. Minsky, N. *et al.* Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature Cell Biol.* **10**, 483–488 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Miyajima, S. Saito and N. Moriyama for experimental support, and M. Yamaki for manuscript preparation. We also thank Y. Maekawa, J. Seta and N. Iwasaki for support with MS. This work was supported in part by The Naito Foundation, the Astellas foundation (to R.F.), the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science (to R.F. and S.K.).

Author Contributions S.K. planned the study with H.K.; R.G.R. and M.B. provided support and general guidance; R.F. designed the study and performed the experiments with H.S. (α -O-GlcNAc purification), A.Y. (LC–MS/MS), W.H. (O-GlcNAc site mapping), T.C. (*in vitro* OGT assay), S.I. (*Drosophila* analysis), Y.I., H.H.H. (ChIP-seq), F.O., J.K. (*in vitro* monoubiquitination assay), K.I. and J.K. (microarray).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K. (uskato@mail.ecc.u-tokyo.ac.jp).

METHODS

Plasmids and retroviruses. Complementary DNAs (cDNAs) of N-terminally Flag-tagged H2B and its mutant were subcloned into pcDNA3 (Invitrogen). A series of H2B point mutants were subcloned into the pET3 vector (Novagen). shRNA sequences targeting hOGT (5'-GCACATAGCAATCTGGCTTCC-3') and *Renilla* luciferase (5'-TGCCTGTAGTACCAAC-3', as a control) were inserted into the pSIREN-RetroQ-ZsGreen vector (Clontech). For retroviral production, the constructed shRNA vectors were transfected into PLAT-A cells. The virus contained in the medium was used for infection.

Generation of stable cell lines. To generate OGT-KD cells by retroviral infection, 10^6 cells were plated in 60 mm culture dishes, treated with 3 ml of retroviral cocktail (1 ml of the prepared retroviral solution plus 2 ml of DMEM with 10% FBS and $8 \mu\text{g ml}^{-1}$ polybrene), then cultured for another 48 h. A FACSVantage (BD) sorter was used to isolate the retrovirally transduced, enhanced green fluorescent protein (eGFP)-positive cells, as previously described⁹. To generate the cells stably expressing Flag-tagged constructs, HeLa cells were transfected with the pcDNA vectors encoding the Flag-tagged H2B or the AA mutant. The cells containing the integrated vectors were selected by exposure to 0.5 mg ml^{-1} G418.

Generation of monoclonal antibody. H2B S112 GlcNAc peptide (CKHAV S(GlcNAc) EGTK) was synthesized (MBL Institute) and used as an antigen (Operon Biotechnologies). The hybridomas were briefly screened using ELISA with the GlcNAc peptide, and finally selected by immunoblot analysis with the *in vitro* GlcNAcylated H2B.

Antibodies. Antibodies were obtained as follows: α -Flag M2 agarose (Sigma), α -H2A, α -H2B, α -H3, α -H4 (Abcam), α -H2B K120 monoubiquitination (Upstate), α -GlcNAc (RL2 or CTD110.6) (Abcam), α -OGT (Sigma), α -Flag (Sigma) and α -RNF20/BRE1A (Bethyl).

Purification and identification of GlcNAc proteins. The α -O-GlcNAc-immobilized beads were prepared with $15 \mu\text{g}$ α -O-GlcNAc (RL2) antibody and 0.5 ml of Dynabeads M-280 sheep α -mouse IgG (Invitrogen) according to the manufacturer's instructions. Chromatin extracts from HeLa cells (0.5 g protein) were prepared essentially as previously described²². In brief, the chromatin pellet, which consisted of residual material from the nuclear extract preparation with buffers supplemented with 1 mM streptozotocin (STZ), was re-suspended with micrococcal nuclease (MNase) buffer (20 mM Tris-HCl, 1 mM CaCl_2 , 2 mM MgCl_2 , 0.1 M KCl, 0.1% (v/v) Triton-X, 0.3 M sucrose, 1 mM DTT, 1 mM benzamidine, 0.2 mM PMSF, 1 mM STZ, pH 7.9). After addition of 3 U ml^{-1} MNase, the samples were incubated for 30 min at room temperature with continuous homogenization and the reaction was stopped by adding 5 mM EGTA and 5 mM EDTA. After centrifugation at $2,000g$ for 30 min at 4°C , the supernatant (chromatin extract) was used for the following purification steps. The chromatin extracts were passed through a WGA agarose column (Vector). The flow-through fraction was further mixed with α -O-GlcNAc-immobilized beads and rotated for 8 h at 4°C . After three washes with buffer D (20 mM Tris-HCl, 0.2 mM EDTA, 5 mM MgCl_2 , 0.1 M KCl, 0.05% (v/v) NP-40, 10% (v/v) glycerol, 1 mM DTT, 1 mM benzamidine, 0.2 mM PMSF, 1 mM STZ, pH 7.9), glycoproteins were eluted twice with buffer D plus 0.4 mg ml^{-1} GlcNAc-O-serine (MBL) (elutions 1 and 2)

and finally with 0.1 M glycine-HCl (pH 2.0) (elution 3). Eluted proteins were desalted by methanol-chloroform precipitation, digested with trypsin (Promega) then loaded on the automated LC-MS/MS system, which was assembled with Zaplous nano-LC (AMR) plumbed with a reverse-phase C18 electrospray ionization (ESI) column (LC assist) and a Finnigan LTQ ion-trap mass spectrometer (Thermo). The LC-MS/MS data were processed using Thermo BioWorks (Thermo) and SEQUEST (Thermo) for protein identification. The list of the identified proteins was further analysed by using the 'gene functional classification tool' in DAVID bioinformatics resources 6.7 (<http://david.abcc.ncifcrf.gov/>).

Recombinant proteins. Preparation of recombinant proteins was performed as previously reported^{9,23}. Recombinant Flag-OGT, Flag-E1, Flag-BRE1A/B complexes were isolated by baculovirus expression and immunoprecipitation-based purification with α -Flag M2 agarose (Sigma). Recombinant $6 \times \text{His}$ -RAD6A was expressed in bacteria and partly isolated with a HIS-Select Nickel Affinity Gel (Sigma). The eluate was diluted 1:20 with BC0 (20 mM HEPES, 0.2 mM EDTA, 10% (v/v) glycerol, pH 7.9), and fractionated with a Resource Q column (GE Healthcare) using a linear gradient (0–0.5 M KCl) method. Preparation of recombinant *Xenopus* histone H2B and its mutants was performed as previously described^{9,23}.

In vitro GlcNAcylation assay (autoradiographic analysis). Recombinant Flag-OGT protein (0.5 μg) was incubated with 0.5 μg of recombinant histone and 0.2 mM (0.2 μCi) UDP-[^3H]GlcNAc (PerkinElmer) in a 25 μl reaction (50 mM Tris-HCl, 12.5 mM MgCl_2 , 1 mM DTT, pH 7.5) for 24 h at 37°C . The reaction was resolved with SDS-PAGE, blotted onto a polyvinylidene difluoride (PVDF) membrane, then subjected to autoradiography after spraying EN³HANCE (NEN Lifescience).

In vitro GlcNAcylation assay (MS analysis). Recombinant histones (1 μg) or recombinant histone octamers assembled *in vitro* (1 μg) were GlcNAcylated by recombinant Flag-OGT in 25 μl reactions (50 mM Tris-HCl, 2 mM UDP-GlcNAc, 12.5 mM MgCl_2 , 1 mM DTT, pH 7.5) for 24 h at 37°C . The reactions were directly subjected to a nano-LC ESI-TOF mass spectrometer system, which was assembled with a 1100 nanoLC (Agilent) plumbed with a ZORBAX 300SB-C18 column (Agilent) and micrOTOF (Bruker). Or, the reactions were digested with trypsin (Promega) and subjected to purification of glycopeptides with an MB-LAC WGA kit (Bruker). The enriched glycopeptides were loaded on the nano-LC ESI-ETD ion-trap mass-spectrometer system, which was assembled with the Agilent HP1200 Nano (Agilent) plumbed with ZORBAX 300SB-C18 (Agilent) and amaZon ETD (Bruker).

In vitro monoubiquitination assay. GlcNAcylated histones (1 μg) were ubiquitinated with the E1 (0.1 μg), RAD6 (0.2 μg), BRE1 complex (0.5 μg), ubiquitin (3 μg) in 50 mM Tris (pH 7.9), 5 mM MgCl_2 , 4 mM ATP at 37°C for 24 h.

ChIP-seq and ChIP-qPCR. ChIP and ChIP-seq libraries were constructed as previously described^{24,25}. For ChIP-seq analysis, the libraries were sequenced to 50 bp with Hiseq2000 (Illumina). For ChIP-qPCR analysis, the fragments of interest in the libraries were quantified with Thermal Cycler TP800 (TAKARA) and SYBR Premix Ex Taq II (Takara). The qPCR primer sets for the GSK3 β gene were 5'-TGCAAGCTCTCAGACGCTAA-3' and 5'-CTCATTTCTCATGGGCGTTT-3'.

An equilibrium-dependent retroviral mRNA switch regulates translational recoding

Brian Houck-Loomis¹, Michael A. Durney², Carolina Salguero², Neelaabh Shankar², Julia M. Nagle², Stephen P. Goff¹ & Victoria M. D'Souza²

Most retroviruses require translational recoding of a viral messenger RNA stop codon to maintain a precise ratio of structural (Gag) and enzymatic (Pol) proteins during virus assembly^{1,2}. Pol is expressed exclusively as a Gag–Pol fusion either by ribosomal frameshifting or by read-through of the *gag* stop codon³. Both of these mechanisms occur infrequently and only affect 5–10% of translating ribosomes, allowing the virus to maintain the critical Gag to Gag–Pol ratio^{4–8}. Although it is understood that the frequency of the recoding event is regulated by *cis* RNA motifs, no mechanistic explanation is currently available for how the critical protein ratio is maintained. Here we present the NMR structure of the murine leukaemia virus recoding signal and show that a protonation-dependent switch occurs to induce the active conformation. The equilibrium is such that at physiological pH the active, read-through permissive conformation is populated at approximately 6%: a level that correlates with *in vivo* protein quantities. The RNA functions by a highly sensitive, chemo-mechanical coupling tuned to ensure an optimal read-through frequency. Similar observations for a frameshifting signal indicate that this novel equilibrium-based mechanism may have a general role in translational recoding.

Genetic recoding either by frameshifting or by read-through requires *cis*-acting elements downstream of the recoding site in the mRNA. In murine leukaemia virus (MLV), a 63-nucleotide read-through signal (Fig. 1a) folds as a pseudoknotted^{7–9} structure (MLV-PK) that can independently direct recoding of stop codons^{5,6,10}, even in heterologous RNA (see Supplementary Text and Supplementary Fig. 1a). Secondary structure mapping indicates that MLV-PK adopts a classic hairpin-type fold⁹ consisting of two stems, S1 and S2, connected by a single base loop L1 and an 18-nucleotide loop L2 (Supplementary Fig. 1a). The 8-base spacer between the UAG stop codon and the first base pair of stem S1 have been proposed to be unstructured^{7,8,11}. Although a majority of RNA pseudoknots contain specific tertiary interactions between stem and loop elements¹², RNA melting data monitored by changes in ultraviolet radiation absorption indicated that these interactions in MLV-PK have an unexpected and pronounced dependence on pH (Supplementary Fig. 2). Assignment of the NMR data (Supplementary Figs 3 and 4) unambiguously confirmed the presence of a pH-dependent equilibrium between two inter-converting conformations. Marked chemical shift perturbations are observed as the minor conformation becomes increasingly populated at lower pH values: this allowed us to select reporter residues, A17 and A14, by which to evaluate the equilibrium (Fig. 1b). As the N1 position of A17 becomes protonated, typical adjacent C2 chemical shift perturbations occur that allow us to calculate the pK_a for this protonation event. Concurrent with protonation at A17, the C2 position of the A14 residue in the minor groove of stem S1 experiences a downfield chemical shift change (Fig. 1b) and we observe the appearance of nuclear Overhauser enhancements (NOEs) from the A14 H2 resonance to ribose resonances of G49 and U50 in loop L2. Together these data show that A14 enters a more electronegative

environment upon the approach of loop L2 and hence indicate the formation of S1–L2 tertiary interactions. Interestingly, curve fitting for both the A17 and A14 resonances yielded a pK_a value of 6.23 and a pH_{fold} value (pH at which 50% of MLV-PK has a tertiary structure) of 6.20, respectively, implying that the two processes of protonation and tertiary structure formation may be coupled (Fig. 1c and Supplementary Fig. 5). We calculate that at physiological pH (7.4), a distribution would result in which ~6% of the MLV-PK population would have a tertiary structure whereas the remaining ~94% would lack the tertiary contacts. As this distribution correlates both with our *in vivo* observations (Supplementary Fig. 6a) and with the previously observed *in vivo* levels of Gag–Pol and Gag^{4–8}, respectively, we asked if an equilibrium ($PK_{active} \rightleftharpoons PK_{inactive}$) determines the Gag:Gag–Pol ratio.

To test functionally this hypothesis, we developed a dual luciferase reporter¹³ *in vitro* translation system and modulated the response by varying the assay pH (Fig. 1d). A control experiment using a scrambled MLV-PK sequence showed that translational read-through by ribosomes is not affected in the pH range tested. These data show that at pH 7.4 ~5.5% of ribosomal recoding occurs—a level in agreement with our measured value of conformational exchange from the NMR experiments. Notably, however, the activity of MLV-PK was found to be strictly dependent upon the pH of the translation system, and exhibits enhanced read-through activity at lower pH (Fig. 1d). Thus, these results indicate that formation of tertiary structure in MLV-PK leads to read-through of the stop codon and, furthermore, that enriching the population with the tertiary fold enhances read-through-stimulating activity. Mutational analysis to test the importance of protonation at A17 uncovered additional protonation sites, which demonstrate that MLV-PK is a complex multiple-proton sensor, and that protonation of A17 is one of two or more chemical triggers that govern the equilibrium (see Supplementary Text and Supplementary Figs 6b and 7).

Although conformational heterogeneity precludes us from assigning all protonation sites and their effect on the global structure of PK_{active} , we can examine the role of A17⁺ by obtaining a partial structure of PK_{active} (see Methods). Comparison with the complete $PK_{inactive}$ structure allows us to propose a mechanism for the conformational transitions that are required for read-through activity (see Supplementary Figs 8, 9, 10 and 11 and Supplementary Table 1). In both conformations a G–U wobble and three distinct A–U Watson–Crick base pairs are observed in the NMR data (Supplementary Fig. 8). Whereas two canonical A–U base pairs are readily assigned to stem S1 (U12–A30 and A14–U27), the third pair can only be formed between bases U6 and A37. Thus, in contrast to previous chemical probing studies⁹, our NMR assignments unambiguously prove that the helical stem S1 is extended by additional base pairing between residues G5–A8 in the spacer and loop L2 residues A34–U38 (Fig. 1a, e and Supplementary Fig. 1). The extended stem S1 helix also contains a 1 × 2 internal bulge^{14,15} formed by bases A8, A34 and G35 (see also Supplementary Fig. 9). Additionally, our structures consistently show residue A29 in

¹Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA. ²Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

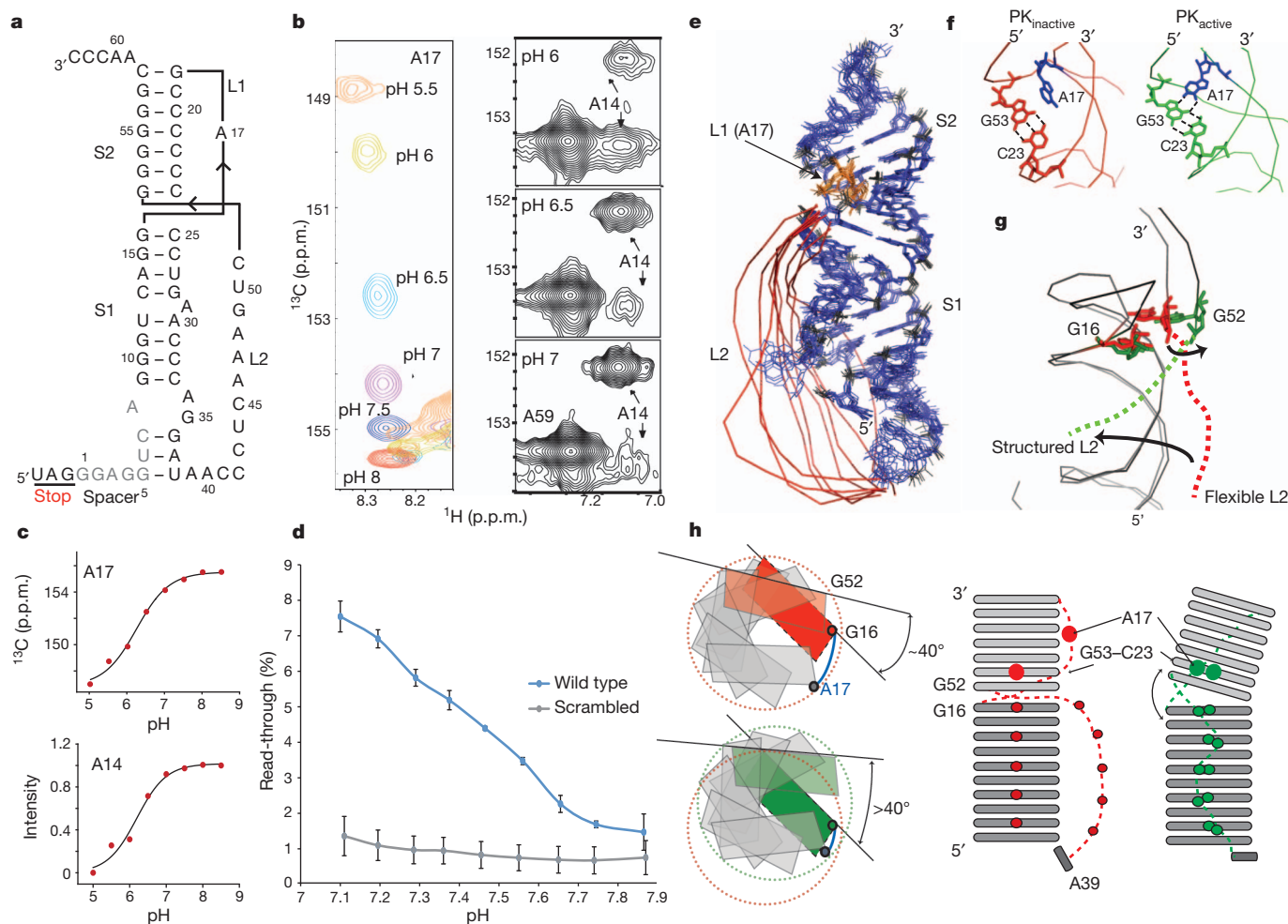


Figure 1 | MLV-PK conformational-equilibrium-dependent read-through. **a**, NMR-derived secondary structure of MLV-PK. Spacer residues (grey) G5–A8 originally predicted to be unstructured are involved in stem S1 formation and consequently loop L2 is shortened. Residue numbers are labelled according to their position in the *in vitro* construct. **b**, Adenosine C2 chemical shift changes in residues A17 and A14 during pH titrations (Supplementary Fig. 5). **c**, Nonlinear data fits²⁶ for residues A17 (top) and A14 (bottom) to determine the pK_a and pH_{fold} , respectively (Supplementary Fig. 5). **d**, The effects of the translation pH on MLV-PK read-through activity *in vitro*. Error bars indicate standard error ($n = 3$). Decreasing the pH from 7.8 to 7.1 leads to a 500% increase in read-through levels. At pH 7.0, our NMR data indicate that ~8% of

MLV-PK is in the folded conformation, which correlates very closely to the *in vitro* read-through level of ~7.5% at pH 7.1. **e**, NMR structure ensemble of $PK_{inactive}$ with residue A17 in orange and the flexible loop L2 in red. A 180° rotated view is shown in Supplementary Fig. 10a. **f**, Close-up views of the stem S2 and loop L1 major groove in both forms of MLV-PK and the A17⁺-C23:G53 base-triple in PK_{active} . Dotted lines indicate hydrogen-bond formation. **g**, Overlay of the $PK_{inactive}$ (red) and PK_{active} (green) structures showing the global transition upon PK_{active} formation. **h**, Schematic of the secondary structure model (right), a view from the helical axis (left). A17 protonation in MLV-PK overwinds the helical twist, which consequently leads to an interhelical bend and releases the steric hindrance at the S1–S2 junction.

stem S1 to be looped out towards the major groove and not involved in tertiary interactions even in PK_{active} (Supplementary Fig. 10). A direct consequence of the 12-base-pair (including the internal A8–G35 stack) extended stem S1 is a shortened 4-nucleotide spacer and a shortened loop L2 that has no long-range NOEs to stem S1 and hence is flexible in $PK_{inactive}$ (Fig. 1e).

In the $PK_{inactive}$ structure (Fig. 1e and Supplementary Table 1 and Supplementary Fig. 10), the extended stem S1 coaxially stacks on stem S2 to form a collinear, quasi-continuous helix as evidenced by typical stacking NOEs between G16 and G52 across the helical junction (Supplementary Fig. 4). Collinear stacking of stems S1 and S2 is possible because S2 contains seven base pairs, which leaves an ~8 Å phosphorus–phosphorus distance for loop L1 across the S2 major groove. This distance can be spanned by the single nucleotide A17 without introducing any major twist or bend at the helical junction¹⁶. However, this coaxial stacking leads to steric hindrance at the junction, which, combined with a shortened loop L2, prevents S1–L2 interactions. In PK_{active} , protonation leads to changes that include a loss of stacking interactions between G16 and G52 at the helical junction, a

marked chemical shift change for residues at the S1–L2 turn (Supplementary Fig. 11) and the appearance of NOEs from G52 H2'/3', G52 H8 and G53 H8 to the protonated A17⁺ H2 (Supplementary Fig. 12). These data indicate that in PK_{active} A17⁺ forms an A17⁺-C23:G53 base triple in the major groove (Fig. 1f). This repositioning of A17⁺ decreases the length of loop L1 below the required minimal distance for coaxial stacking and, accordingly, an inter-helical bend is introduced that relieves the steric hindrance at the junction and allows for S1–L2 tertiary interactions (Figs 1f–h).

Because our data show that MLV-PK contains critical motifs that may regulate the structural transition (A17, A29, the helical junction and the S1–L2 turn), we proposed that manipulating these sites to favour tertiary contacts may modulate the equilibrium between PK_{active} and $PK_{inactive}$. Given that adenosine-stacking interactions are known to increase thermodynamic stability at turns¹⁷, we asked if a U38A mutant would allow for increased A38–A39 stacking at the S1–L2 turn and explain the previously observed increase in read-through levels⁸. In addition, we used an A29C mutant to test if the looped-out A29 sterically hinders S1–L2 triplex formation.

Read-through levels *in vivo* are significantly stimulated in both U38A (~80% increase) and A29C (~20%) (Supplementary Fig. 6c). Remarkably, the pK_a for A17 in the U38A mutant is increased to 6.40 (Fig. 2a, b) and inter-helical bending occurs at higher pH (data not shown). Consequently, this construct has an increased response to pH change (Fig. 2a). These data indicate that improved adenosine stacking in U38A facilitates S1–L2 folding, which is tightly coupled to the protonation event at the distal A17 site via the helical junction. Next, we asked if the two hyperactive mutants could have an additive effect in a U38A:A29C double mutant. Notably, a synergistic effect was observed for the double mutant, affording a ~260% increase in read-through levels *in vivo*, which correlates with a significantly elevated pK_a of 7.09 for A17 (Fig. 2b, c and Supplementary Fig. 13). Furthermore, we could also design mutants that disrupt the structural integrity of the stem S1 and the S1–L2 turn and shift the equilibrium towards the $PK_{inactive}$ form: the pH response is thus attenuated (G15A:C26U) or abrogated (G11C, U6A, A39U) (Fig. 2a and Supplementary Figs 6c and 11b).

Next, we decided to engineer an artificial inter-helical bend by deleting one base pair in S2 (ΔGC) to test the importance of the mechanical response. Noticeably, our NMR data for ΔGC show that although the junction is unstacked at high pH, its ability to form the PK_{active} conformer is abrogated (Fig. 3a and Supplementary Fig. 14). This construct is not pH responsive, indicating that the correct register for the S1–L2 triple-helix formation does not form. We were able to rescue this mutant by using a ΔGC :U38A double mutant (Fig. 3b). Because U38A stabilizes the S1–L2 turn, this construct readjusts the inter-helical bend by bringing loop L2 in register with stem S1. This restores the pH response and conclusively proves that long-range effects occur between the junction and the S1–L2 turn. Although it was not possible to perform the *in vitro* translation experiment on the ΔGC mutant owing to the associated introduction of an in-frame U38AA stop codon, we could test the ΔGC :U38A construct (Fig. 3b). In agreement with our NMR data this construct rescues the pH-modulated recoding activity, albeit with a lower efficiency compared to U38A (Fig. 3b). These experiments prove that the mechanical and chemical responses in MLV-PK are indeed coupled, and confirm our equilibrium-based RNA conformational switch model. Finally, similar pH-dependent, frameshift activity by the well-characterized frameshifting signal from beet western yellows virus (BWYV)^{18,19} pseudoknot (Fig. 3c) suggests that fine-tuning a proton-driven structural equilibrium may be a common mechanism for regulating recoding of gene expression.

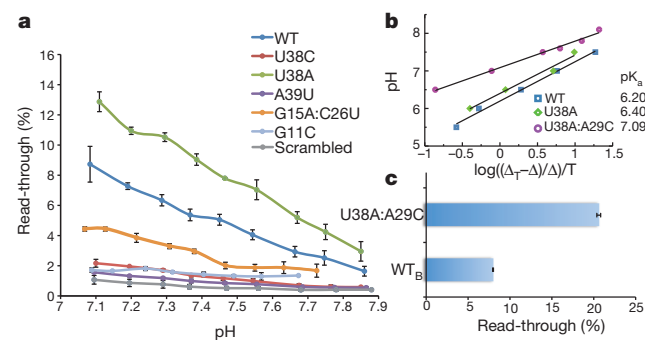


Figure 2 | Functional data for modulation of the equilibrium. **a**, The effects of the translation pH on read-through activity of MLV-PK S1–L2 turn mutants *in vitro*. The data show the positive effect of increased stacking in U38A, which is abrogated in U38C and correlates with *in vivo* activity (Supplementary Fig. 6c). Error bars indicate standard error ($n = 3$). **b**, Hill plots²⁷ ($R \geq 0.99$) to determine the pK_a values of A17 in the indicated MLV-PK constructs. Δ_T is the total chemical shift difference for the C2 carbon of A17 in $PK_{inactive}$ and PK_{active} and Δ is the difference between the C2 chemical shift at a given pH and the maximum chemical shift. **c**, *In vivo* read-through levels for the MLV-PK wild type (WT_B) and the U38A:A29C mutant.

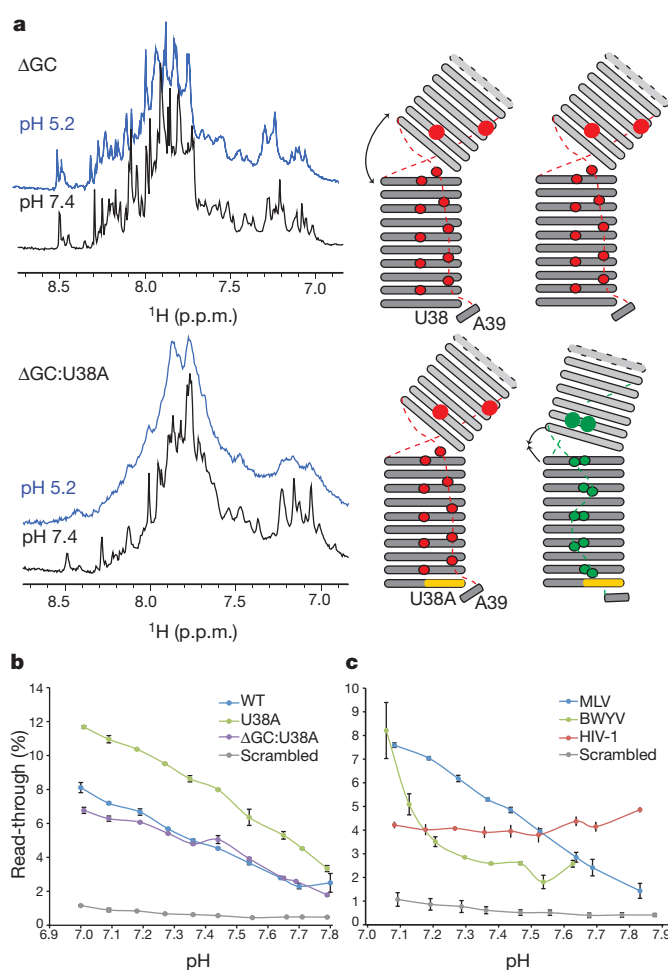


Figure 3 | Compensatory effects of the S1–L2 turn on the engineered inter-helical bend and equilibrium-based frameshifting levels. **a**, One-dimensional NMR data (left) at pH 7.4 (black) and pH 5.2 (blue) for the indicated MLV-PK constructs to probe the long-range conformational transition. See Supplementary Fig. 2b for comparison with wild-type MLV-PK. Corresponding schematic topology models (right) indicate the degree of inter-helical bending and compensation by the S1–L2 turn. Broadening of the line widths are indicative of tertiary structure formation, which is abrogated in ΔGC but compensated in the ΔGC :U38A mutant. **b**, The effects of the translation pH on the *in vitro* read-through activity of the above MLV-PK constructs. Error bars indicate standard error ($n = 3$). **c**, Comparison of the effects of the translation pH on MLV-PK read-through and BWYV and human immunodeficiency virus-1 (HIV-1) frameshift activity *in vitro*. Error bars indicate standard error ($n = 3$). In HIV-1, the recoding signal is understood to be a stem-loop²⁸ in which no protonation sites are predicted. In our pH-modulated assay the HIV-1 construct is not dependent on protonation for activity.

RNA is capable of a wide range of structural transitions ranging from subtle rearrangements to global folding that determine function²⁰. MLV-PK is a highly responsive, multi-proton sensor that couples a chemical and mechanical response to induce a conformational transition. We present a novel equilibrium-based mechanism for regulation of the frequency of ribosomal recoding, wherein the only determinant required is an RNA conformational switch (Fig. 4). Previous explanations for how the recoding frequency is maintained have invoked mechanisms such as ribosomal mechanical stress, thermodynamic and kinetic stability of the recoding signal, ribosomal pausing and ribosomal helicase unwinding^{21,22}. These processes will have a role in recoding but, according to our proposed mechanism, they would occur only if the translating ribosome encounters the active conformation of the RNA. Further support for our mechanism comes from the slow exchange lifetime of MLV-PK on the NMR time scale (>100 ms), which is slower than the known rate of ribosomal decoding

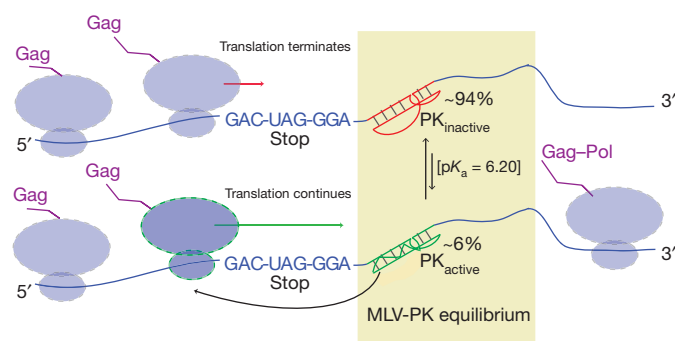


Figure 4 | Model for equilibrium-based mechanism. Gene expression via translational recoding is regulated by a dynamic equilibrium ($PK_{active} \rightleftharpoons PK_{inactive}$) between an active, read-through permissive conformation and an inactive, non-permissive conformation the distribution of which determines the Gag:Gag-Pol ratio. Ribosomes that encounter the PK_{active} conformation continue translating through the stop codon. Structural and functional data indicate that PK_{active} interacts with the ribosome before the stop codon is in the decoding centre owing to the shortened nature of the linker.

*in vivo*²³. This ensures that the pseudoknot tertiary fold will be maintained while a translating ribosome attempts to decode the Gag:Gag-Pol boundary.

METHODS SUMMARY

Detailed methods for RNA sample preparation, biophysical characterization^{12,24}, structure determination^{24,25} and translational recoding assays¹³ can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 May; accepted 20 October 2011.

Published online 27 November 2011.

1. Felsenstein, K. M. & Goff, S. P. Expression of the Gag-Pol fusion protein of Moloney murine leukemia virus without Gag protein does not induce virion formation or proteolytic processing. *J. Virol.* **62**, 2179–2182 (1988).
2. Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.* **75**, 1834–1841 (2001).
3. Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Recoding: translational bifurcations in gene expression. *Gene* **286**, 187–201 (2002).
4. Philipson, L. *et al.* Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the Gag-Pol polypeptide with yeast suppressor tRNA. *Cell* **13**, 189–199 (1978).
5. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl Acad. Sci. USA* **82**, 1618–1622 (1985).
6. Wills, N. M., Gesteland, R. F. & Atkins, J. F. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus *gag* stop codon. *Proc. Natl Acad. Sci. USA* **88**, 6991–6995 (1991).
7. Wills, N. M., Gesteland, R. F. & Atkins, J. F. Pseudoknot-dependent read-through of retroviral *gag* termination codons: importance of sequences in the spacer and loop 2. *EMBO J.* **13**, 4137–4144 (1994).
8. Alam, S. L., Wills, N. M., Ingram, J. A., Atkins, J. F. & Gesteland, R. F. Structural studies of the RNA pseudoknot required for readthrough of the *gag*-termination codon of murine leukemia virus. *J. Mol. Biol.* **288**, 837–852 (1999).
9. Brierley, I., Pennell, S. & Gilbert, R. J. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Rev. Microbiol.* **5**, 598–610 (2007).

10. Honigman, A., Wolf, D., Yaish, S., Falk, H. & Panet, A. cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology* **183**, 313–319 (1991).
11. Feng, Y. X., Yuan, H., Rein, A. & Levin, J. G. Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the *gag* termination codon followed by an RNA pseudoknot. *J. Virol.* **66**, 5127–5132 (1992).
12. Nixon, P. L. & Giedroc, D. P. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.* **296**, 659–671 (2000).
13. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479–486 (1998).
14. Schroeder, S., Kim, J. & Turner, D. H. G. A. and U.U mismatches can stabilize RNA internal loops of three nucleotides. *Biochemistry* **35**, 16105–16109 (1996).
15. Badhwar, J., Karri, S., Cass, C. K., Wunderlich, E. L. & Znosko, B. M. Thermodynamic characterization of RNA duplexes containing naturally occurring 1 × 2 nucleotide internal loops. *Biochemistry* **46**, 14715–14724 (2007).
16. Michiels, P. J. *et al.* Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**, 1109–1123 (2001).
17. Clanton-Arrowood, K., McGurk, J. & Schroeder, S. J. 3' terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices. *Biochemistry* **47**, 13418–13427 (2008).
18. Cornish, P. V., Hennig, M. & Giedroc, D. P. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated –1 ribosomal frameshifting. *Proc. Natl Acad. Sci. USA* **102**, 12694–12699 (2005).
19. Atkins, J. F. & Gesteland, R. F. *Recoding: Expansion of Decoding Rules Enriches Gene Expression* (Springer, 2010).
20. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).
21. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
22. Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* **139**, 193–208 (2009).
23. Agirrezabala, X. & Frank, J. Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Q. Rev. Biophys.* **42**, 159–200 (2009).
24. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* **337**, 427–442 (2004).
25. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
26. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* **38**, 4896–4903 (1999).
27. Legault, P. & Pardi, A. *In-situ* probing of adenine protonation in RNA by ^{13}C NMR. *J. Am. Chem. Soc.* **116**, 8390–8391 (1994).
28. Staple, D. W. & Butcher, S. E. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J. Mol. Biol.* **349**, 1011–1023 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Atkins for the p2luc plasmid, the New York Structural Biology Center for NMR time and A. Hawkins, S. Leiman and J. Gullinger for their contributions. B.H.-L. and S.P.G. would also like to thank the late D. Wolf for his inspiration and many helpful discussions. S.P.G. acknowledges grant R37 CA30488 from the NCI/NIH and is an Investigator with the Howard Hughes Medical Institute.

Author Contributions V.M.D'S., B.H.-L., M.A.D. and S.P.G. conceived of and designed the experiments. V.M.D'S., M.A.D., C.S., N.S. and B.H.-L. did the structural analysis, B.H.-L. did the recoding assays, and M.A.D. and J.M.N. did the *in vivo* assay for the double mutant. M.A.D. and V.M.D'S. wrote the manuscript with assistance from B.H.-L.

Author Information Coordinates and restraints for the final ensemble of ten structures of $PK_{inactive}$ have been deposited in the Protein Data Bank under accession code 2LCS. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.M.D'S. (dsouza@mcb.harvard.edu).

METHODS

RNA sample preparation. DNA sequences were designed to include the T7 promoter, an insert sequence corresponding to nucleotides 1–63 of MLV-PK, a SmaI linearization site and restriction sites for BamHI and EcoRI for insertion of the amplified product into pUC19. RNA samples for biophysical experiments were transcribed and purified as described²⁴. Samples were prepared in various buffers as required for the experiments (see later).

UV melting analysis. RNA samples for the wild-type, U38A and A39U MLV-PK constructs were annealed at 95 °C under dilute conditions (1–2 µM) in UV buffer (20 mM cacodylate, 1 M NaCl). Absorbance versus temperature melting curves for the MLV-PK RNA samples were acquired at wavelengths of 260 nm and 280 nm. The heating rate was fixed at 0.5 °C min^{−1} with a Beckman DU800 single beam spectrophotometer equipped with a Peltier heating device. The pH experiments were performed at pH 5.5 and 7.5 and absorbance data were converted to first derivative versus temperature plots and analysed as described previously¹².

In vivo translation assay. Read-through was monitored using the p2luc dual luciferase reporter system¹³. The complete MLV-PK, BWYV and HIV-1 recoding sequence (MLV-PK, GACCCTAGATGACTAGGGAGGTCAGGGTCAGGAG CCCCCCCTGAACCCAGGAAACCCCTCAAAGTCGGGGGGCAACCCGTC; BWYV, CAATTCATCGGGAACTAAGTGC GCGGCACCGTCCGCGGAAC AAACGGAAG; HIV-1, GAGACAGGCTAATTTTCTTACAGAGCAGACCATAGCC) were inserted between the *Renilla* (luc) and firefly (luc) luciferase genes using either the SmaI and BamHI or the BamHI and SacI restriction sites. To normalize for transfection and translation efficiency, control constructs were transfected in parallel to the sample plasmid. The control plasmid was identical to the MLV-PK construct except for a single U to C mutation in the stop codon and represented a read-through efficiency of 100% at the mutated codon. The control constructs for BWYV and HIV-1 contained the same sequences in the −1 reading frame and the slippery sites were abrogated: (BWYV, CAATTCATCGGGAACTAAGTGC GCGGCACCGTCCGCGGAACAAACGGAAG; HIV-1, GAGACAGGCTAATTT CTTAAGGGAAGATCTGGCCTTCCACAAGGGAAGGCCAGGGAATTTT CTTACAGAGCAGACCATAGCC). Dual luciferase assays were performed in 96-well format. 293A cells (Invitrogen) were plated at a density of 2×10^4 cells per well in 125 µl DMEM plus penicillin and streptomycin and 10% FBS and grown overnight at 37 °C in 5% CO₂. Transfections were performed with Fugene6 (Roche) and transfection mixes were made according to the manufacturer's instructions (50 ng reporter, 0.150 µl Fugene6, serum-free medium in a 17.5 µl final volume). Transfections were performed in triplicate and 5 µl of the transfection mix was added to each well. Twenty-four hours post-transfection, the supernatant was decanted and 25 µl 1× PLB (passive lysis buffer; Promega) was added to each well. Lysates were incubated at ambient temperature with agitation

for 10 min then subjected to one freeze–thaw cycle (−80 °C). Dual luciferase measurements were performed with 20 µl of the lysate using a dual luciferase reporter assay system (Promega). The luciferase values from the control were used to normalize those from the sample, allowing us to report read-through normalized to 100%.

In vitro translation assay. Dual luciferase plasmids containing the MLV-PK recoding sequence inserted between the SmaI and BamHI sites were linearized with HpaI and recovered by ethanol precipitation. The linearized template (1 µg) was used to synthesize capped RNA using the mMESSAGE mMACHINE kit (Ambion) and the manufacturer's protocol (2 h incubation). The capped RNA was polyadenylated for 1 h using the poly(A) tailing kit (Ambion) and the mRNA was purified using the MEGAClear kit (Ambion). RNA integrity was confirmed by running 1 µg on a denaturing glyoxal gel. The pH of the rabbit reticulocyte lysate (Ambion) was altered by the addition of dilute HCl and NaOH, the concentration of which was empirically determined. 5.75 µl of HCl (9.35, 6.32, 3.33, 0.0 mM) or NaOH (0.53, 1.30, 2.04, 2.78, 3.51, 4.23, 4.93, 5.63, 6.34 mM) was dispensed into 0.2 ml PCR tubes. Translation mixes (prepared in triplicate) were combined on ice and contained 1.25 µl 20× translation buffer (Ambion), 0.5 µl 50× (2.5 mM) methionine, 17 µl lysate and 0.5 µl (0.125 µg) reporter mRNA. The translation mix was added to the dilute acid or base, briefly vortexed, and incubated at 30 °C for 1.5 h. The reaction was quenched on ice for 10 min, 5 µl was removed for the dual luciferase assay and the pH of the remaining reaction was measured at 30 °C with an InLab Micro pH probe (Mettler-Toledo).

NMR data acquisition, resonance assignment and structure calculations. For NMR experiments RNA samples were resuspended in NMR buffer (10 mM Tris-HCl at pH 6.5 and pH 7.5 for PK_{active} and PK_{inactive}, respectively, and 10 mM NaCl). NMR data were acquired using Bruker 700 MHz and 900 MHz spectrometers equipped with cryoprobes. Spectra were recorded at 298 K and 308 K with the exception of data for the imino region for which data were also recorded at 278 K. Assignments for non-exchangeable ¹H and ¹³C signals were obtained from two-dimensional NOESY, two-dimensional HMQC and three-dimensional HMQC-NOESY data sets recorded with unlabelled and nucleotide-specific (AC^{CN}-MLV-PK, GC^{CN}-MLV-PK) selectively labelled samples and two-dimensional NOESY samples obtained for nucleotide-specifically deuterated samples (GU^H-MLV-PK and AC^H-MLV-PK)²⁴. Structures were calculated as described²⁴ using manually assigned restraints in CYANA²⁵. The statistics table for the PK_{inactive} structure ensemble is included in Supplementary Table 1. The changes that occur upon A17 protonation in stem S2, loop L1 and the helical junction were used to generate a model for this region of PK_{active}. The NOEs were confirmed with various MLV-PK mutants (U6A, A17U, A17C) (see Supplementary Figs 11 and 12). Molecular images were generated with PyMOL (<http://www.pymol.org>).

Structures of the multidrug exporter AcrB reveal a proximal multisite drug-binding pocket

Ryosuke Nakashima^{1*}, Keisuke Sakurai^{1*}, Seiji Yamasaki², Kunihiro Nishino³ & Akihito Yamaguchi^{1,2}

AcrB and its homologues are the principal multidrug transporters in Gram-negative bacteria^{1–6} and are important in antibiotic drug tolerance^{7,8}. AcrB is a homotrimer that acts as a tripartite complex^{9,10} with the outer membrane channel TolC^{11,12} and the membrane fusion protein AcrA^{13,14}. Minocycline and doxorubicin have been shown to bind to the phenylalanine cluster region of the binding monomer¹⁵. Here we report the crystal structures of AcrB bound to the high-molecular-mass drugs rifampicin and erythromycin. These drugs bind to the access monomer, and the binding sites are located in the proximal multisite binding pocket, which is separated from the phenylalanine cluster region (distal pocket) by the Phe-617 loop. Our structures indicate that there are two discrete multisite binding pockets along the intramolecular channel. High-molecular-mass drugs first bind to the proximal pocket in the access state and are then forced into the distal pocket in the binding state by a peristaltic mechanism involving sub-domain movements that include a shift of the Phe-617 loop. By contrast, low-molecular-mass drugs, such as minocycline and doxorubicin, travel through the proximal pocket without specific binding and immediately bind to the distal pocket. The presence of two discrete, high-volume multisite binding pockets contributes to the remarkably broad substrate recognition of AcrB.

The crystal structure of the AcrB trimer¹⁶ indicates that drugs are taken up from the vestibules on the side of AcrB and are subsequently released from the funnel-like opening at the top of the AcrB trimer into the TolC channel. Our first AcrB crystal had crystallographic three-fold symmetry¹⁶ (space group, R32). Although there have been reports of symmetrical crystal structures of AcrB bound to drugs and to substrates^{17–20}, neither we nor another group²¹ detected any binding to drugs for the R32 crystal of AcrB. Subsequently, we obtained a different AcrB crystal that did not have crystallographic three-fold symmetry (space group, C2) and determined a structure for AcrB bound to minocycline and doxorubicin using this crystal¹⁵. Notably, for each drug only one molecule was bound to the AcrB trimer. Each of the three AcrB monomers represents a different intermediate state of the transport reaction: ‘access’, ‘binding’ and ‘extrusion’. Drugs bind only to the binding monomer, and the binding sites are located in the phenylalanine cluster region¹⁵. Intramolecular channels connect the vestibule to the drug-binding site, but the exit from the binding site to the funnel-like opening at the top of the AcrB trimer is closed (inside-open form). By contrast, the extrusion monomer has an open exit and a smaller vacant binding site, and the channel from the vestibule is closed (outside-open form). The access monomer has an inside-open structure, but also has a smaller binding site. Drugs are transported by a functionally rotating mechanism^{15,22} that is also found in native, asymmetric crystals of AcrB^{23–25} and its close homologue MexB²⁶.

The binding sites for minocycline (relative molecular mass, $M_r \sim 457$) and doxorubicin ($M_r \sim 544$) only partly overlap, and this type of multisite binding has been reported previously for the transcriptional regulator QacR²⁷. However, AcrB has a much broader

substrate specificity¹ than QacR. In this study, we determined the crystal structure of AcrB in complex with the high-molecular-mass drugs (HMMDs) rifampicin ($M_r \sim 823$) and erythromycin ($M_r \sim 734$). As a result, we discovered an additional multisite binding pocket (proximal binding pocket) that is separated from the pocket where minocycline and doxorubicin bind (distal binding pocket).

The crystal structure of rifampicin-bound AcrB is generally consistent with previous structures determined for minocycline-bound and doxorubicin-bound AcrB¹⁵ (Fig. 1a and Supplementary Fig. 2a). In the difference Fourier map that we constructed between rifampicin-bound and drug-free AcrB at 4.0σ , we detected electron density for only one drug molecule bound to the trimer (Fig. 1a and Supplementary Fig. 2a, b). The rifampicin (Supplementary Fig. 1a) molecule is bound to the access monomer and not the binding monomer (Fig. 1b) and there is no extra electron density observed at the corresponding region in the binding or extrusion monomers.

Rifampicin is sandwiched between the β -sheets of the PC1 and PC2 subdomains in the access monomer. Arg 717 of the PC2 subdomain is hydrogen-bonded to the oxygen atom of the 25-acetoxy group of rifampicin and Asn 719 is hydrogen-bonded to the 1-hydroxyl group of the aromatic ring and the 21-hydroxyl group of the *ansa*-chain (Fig. 1c and Supplementary Figs 1 and 2c). Phe 617 of the PC1 subdomain interacts hydrophobically with the 12-methyl group of rifampicin.

Phe 617 is located at the tip of the hairpin-like loop between the C β 2' and C β 3' chains of PC1 (Fig. 1c and Supplementary Fig. 2c). We noticed this distinctive structure in our previous report¹⁶ and referred to it as a ‘hook-like bend’; however, we did not know its significance at that time. Phe 617 also interacts with doxorubicin in the binding monomer of doxorubicin-bound AcrB¹⁵. The rifampicin- and doxorubicin-binding sites are located back to back across the Phe-617 loop (Supplementary Fig. 3), which forms a partition between the proximal and distal binding pockets. Furthermore, there is a narrow channel from the proximal binding pocket to the distal binding pocket under the Phe-617 loop (Supplementary Fig. 3). In the binding monomer, the Phe-617 loop swings towards the rifampicin-binding site and, as a result, prevents rifampicin from binding to the proximal pocket in the binding monomer (Supplementary Movie 4a, b).

We determined the structure of erythromycin-bound AcrB (Fig. 2a and Supplementary Fig. 5a) and found that erythromycin (Supplementary Fig. S1) binds to the access monomer. The erythromycin-binding site partly overlaps the rifampicin-binding site. When the two binding forms are superimposed (Fig. 2b and Supplementary Fig. 5b), the macrolide ring of erythromycin is located under rifampicin and is perpendicular to its macrocyclic ring.

In the erythromycin-bound structure, the Phe-617 loop is located farther towards the rifampicin-binding site (Supplementary Fig. 6) than in the rifampicin-bound structure. The macrocyclic ring of erythromycin binds in the space under the Phe-617 loop, towards the phenylalanine cluster region, and partly overlaps the doxorubicin-binding site. This

¹Department of Cell Membrane Biology, Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567-0047, Japan. ²Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka 565-0871, Japan. ³Laboratory of Microbiology and Infectious Diseases, Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka 567-0047, Japan.

*These authors contributed equally to this work.

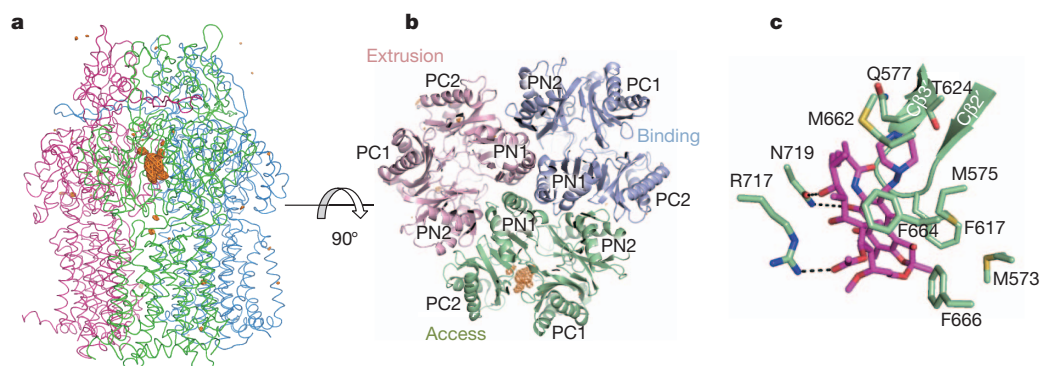


Figure 1 | Crystal structure of the rifampicin-bound AcrB trimer. The three AcrB monomers are shown in blue, red and green to indicate the binding, extrusion and access monomers, respectively. This colour scheme is used in all the figures. **a**, Entire structure of the AcrB trimer with rifampicin viewed from the side and parallel to the membrane plane. The difference Fourier map ($F_{\text{drug}} - F_{\text{free}}$) of bound rifampicin is depicted by an orange mesh, contoured at

4.0 σ . **b**, Cutaway view of the head piece of the AcrB trimer from the distal side of the cell. **c**, Close-up view of the rifampicin-binding site. Carbon atoms of rifampicin and AcrB are shown in magenta and green, respectively. Nitrogen, oxygen and sulphur atoms are shown in blue, red and yellow, respectively. Hydrogen bonds are indicated by dotted lines.

ring is sandwiched between Phe 617 and the loop containing Ser 134 and Ser 135. Additionally, the side chains containing residues Ser 79 and Thr 91 in the PN1 subdomain and the side chain containing residue Asp 681 in the PC2 subdomain surround the desosamine moiety of erythromycin, and Leu 674 is located near the cladinose moiety (Fig. 2c and Supplementary Fig. 5c).

In the binding monomer, the distance between Phe 617 and the pair of serine residues Ser 134 and Ser 135 is greater than in the access monomer, and as a result the erythromycin molecule does not fit into the proximal binding site in the binding monomer (Supplementary Movie 7a and Supplementary Fig. 7b). The distal pocket in the binding monomer is large enough to accommodate erythromycin, but we did not detect this drug in the distal pocket. The erythromycin is probably absent because the path under the Phe-617 loop is too narrow for the drug (as it is for rifampicin) to pass through in the purified proteins. During the dynamic drug efflux process, erythromycin and rifampicin would be forced through this path by a transient conformational change from the access form to the binding form, and the drugs would be present in the distal pocket in the binding state.

As shown in Fig. 3, rifampicin and minocycline bind to the AcrB trimer simultaneously; these drugs bind to the access and binding monomers, respectively, as indicated by the difference Fourier map ($F_{\text{drug}} - F_{\text{free}}$) at 4.0 σ (Supplementary Movie 8).

To clarify the *in vivo* roles of the proximal and distal binding pockets of AcrB, we constructed a series of site-directed mutants (Fig. 4a and Supplementary Fig. 9a, b). The expression level of AcrB was not

significantly affected by these mutations (Supplementary Fig. 10), and the effects of the mutations on the erythromycin (proximal-binding drug) and doxorubicin (distal-binding drug) resistances of *Escherichia coli* cells expressing mutant AcrB were determined (Fig. 4b and Supplementary Fig. 11a, respectively).

The drug export activity of AcrB was measured by its ability to prevent the accumulation of doxorubicin in *E. coli* cells. As shown in Fig. 4c, *acrB*-deficient *E. coli* cells accumulated doxorubicin, and doxorubicin fluorescence in the medium showed time-dependent quenching. By contrast, AcrB-expressing cells did not show quenching of doxorubicin fluorescence because AcrB immediately expelled the doxorubicin that was incorporated into the cell. The addition of minocycline (a low-molecular-mass drug (LMMD)), erythromycin or rifampicin (both HMMDs) inhibited doxorubicin extrusion and restored the decrease in doxorubicin fluorescence in AcrB-expressing cells, indicating that LMMDs and HMMDs use a common translocation pathway. Furthermore, kanamycin did not restore doxorubicin fluorescence quenching because AcrB does not export kanamycin. Additionally, doxorubicin fluorescence itself was not affected by these antibiotics (Supplementary Fig. 12).

Ser 134 and Ser 135 form hydrogen bonds with erythromycin (Fig. 2c and Supplementary Fig. 5c). Residues Ser 79 and Thr 91 do not directly interact with erythromycin (Fig. 4a), but the introduction of long side chains at these positions would prevent erythromycin binding. Asp 681 is located in the deep interior of the proximal pocket, and the introduction of a high-volume side chain at this position also

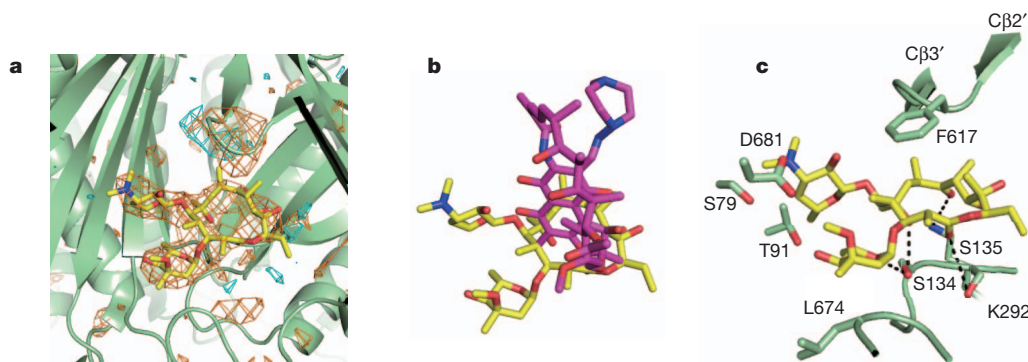


Figure 2 | Crystal structure of the erythromycin-binding site of AcrB with a bound erythromycin molecule. Carbon atoms of erythromycin are shown in yellow. The other colours indicate the same as in Fig. 1c. **a**, Close-up view of the erythromycin-binding site. Bound erythromycin is shown in yellow, and the difference Fourier map with positive peaks (orange mesh, contoured at 3.0 σ)

and negative peaks (cyan mesh, contoured at -3.5σ) is shown. **b**, Overlapping structures of rifampicin and erythromycin at the binding site of AcrB. **c**, Erythromycin binding site of AcrB with a bound erythromycin molecule. Hydrogen bonds are indicated by dotted lines.

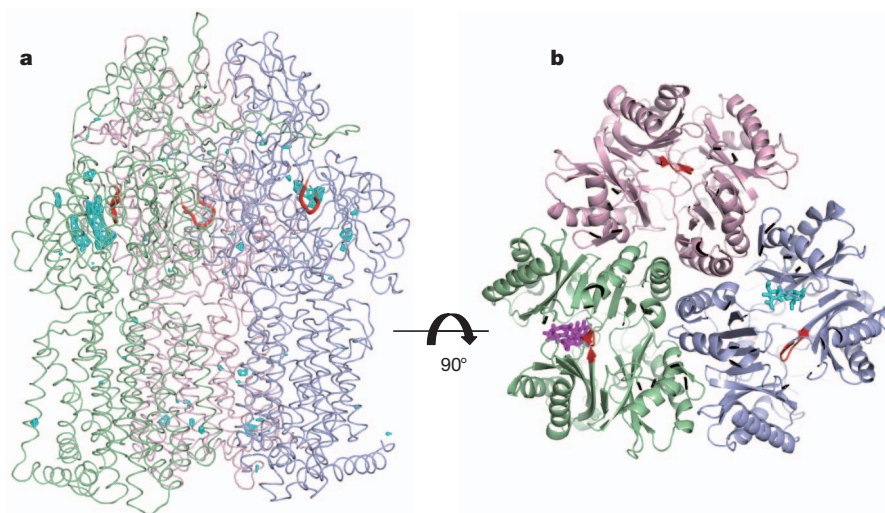


Figure 3 | Structure of the AcrB trimer with simultaneously bound rifampicin and minocycline. **a**, Side view of AcrB with a difference Fourier map of bound rifampicin in the access monomer and minocycline in the binding monomer, which is depicted by a cyan mesh, contoured at 4.0σ .

b, Horizontal cutaway view of AcrB. Rifampicin and minocycline are shown in magenta and cyan, respectively, using stick representations, and the Phe-617 loops are shown in red.

would prevent erythromycin binding (Fig. 4a). The double mutations Ser134Ala Ser135Ala and Ser79Arg Thr91Arg and the single mutation Asp681Trp all significantly decreased cell growth at subinhibitory concentrations ($32\mu\text{g ml}^{-1}$) of erythromycin (Fig. 4b). However, doxorubicin resistance (Supplementary Fig. 11a) and its efflux transport (Fig. 4d) were not affected, probably because of the weak interaction of doxorubicin with the proximal pocket and because of the space available for doxorubicin to travel through the pocket. In these proximal mutants, the inhibition of doxorubicin efflux by erythromycin completely disappeared (Fig. 4e), supporting the idea that the drugs

that bind proximally and distally pass through a common translocation pathway.

The distal pocket mutations Phe610Ala and Phe628Ala reduced both doxorubicin and erythromycin resistance to varying degrees (Fig. 4b and Supplementary Fig. 11a), which indicates that both drugs travel through the distal pocket *in vivo*. Additionally, doxorubicin efflux activity was completely abrogated in these distal mutants (Fig. 4d).

The drug-binding AcrB structure revealed that the Phe-617 loop separates the proximal and distal pockets, and that the orientation of

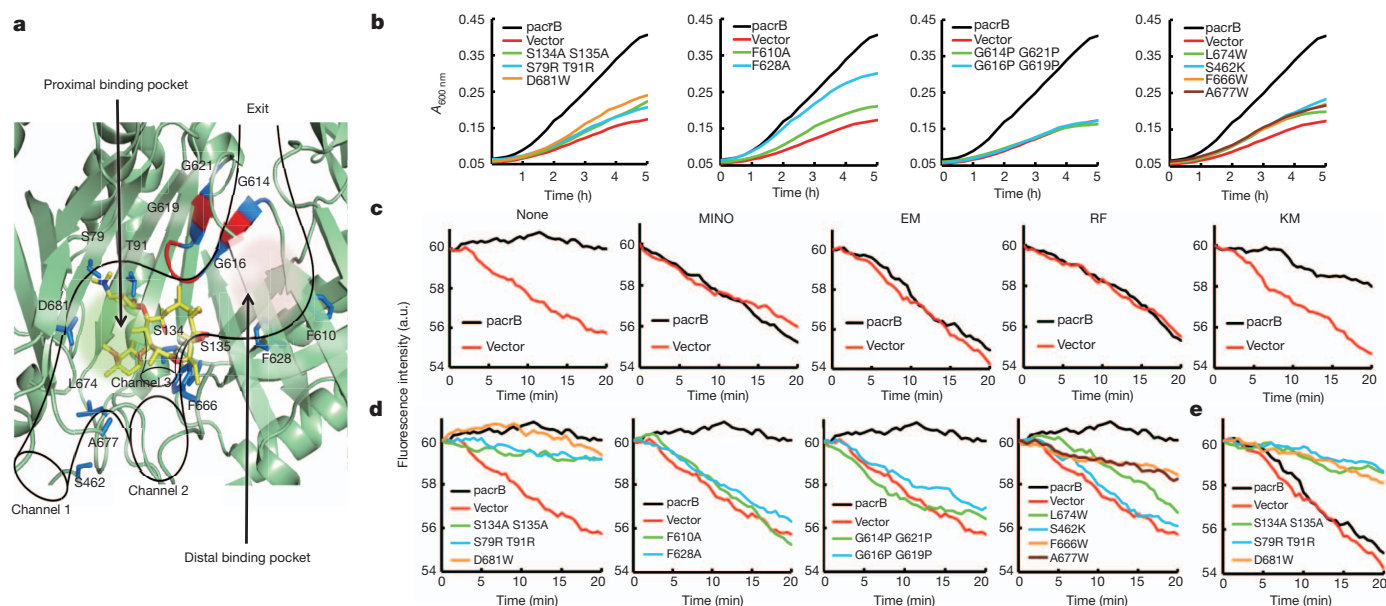


Figure 4 | Effect of site-directed mutagenesis in the two binding pockets and putative intramolecular channels. **a**, Close-up view of the erythromycin-binding site with mutated amino-acid residues indicated by blue sticks. The branched intramolecular channels are outlined in black, the Phe-617 loop (Gly 614–Gly 621) is shown in red and bound erythromycin is shown in yellow. For clarity, some residues have been removed from the foreground. **b**, Growth of mutant-AcrB-expressing *E. coli* cells in the presence of $32\mu\text{g ml}^{-1}$

erythromycin. **c–e**, Quenching of doxorubicin fluorescence as a result of doxorubicin accumulation in intact mutant-AcrB-expressing *E. coli* cells: competitive inhibition of doxorubicin export by various drugs (MINO, minocycline; EM, erythromycin; RF, rifampicin; KM, kanamycin) (**c**); effect of AcrB mutations on doxorubicin export (**d**); effect of erythromycin on doxorubicin export by mutant AcrB (**e**). a.u., arbitrary units.

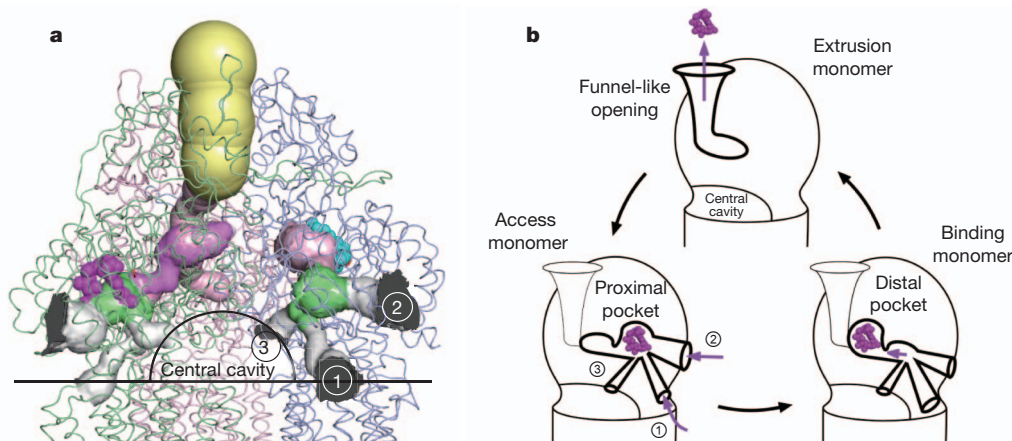


Figure 5 | Crystal structure of the rifampicin–minocycline-bound AcrB trimer. **a**, Side view of the AcrB trimer with intramolecular channels and bound drugs. The channels are shown as coloured solid surfaces and were calculated using the program CAVER²⁹; the channels include the proximal binding pocket (light green), the distal binding pocket (light pink) and the exit funnel (light yellow). Three channels are labelled, and bound drugs are shown

this loop alters the active binding pockets (Supplementary Fig. 3 and Supplementary Movie 4). However, the Phe617Ala mutation has been reported to have no effect on resistance to erythromycin or doxorubicin²⁸. Therefore, we introduced the double mutations Gly614Pro Gly621Pro and Gly616Pro Gly619Pro to prevent movement of the Phe-617 loop. The growth of these mutants under a subinhibitory concentration of either erythromycin (Fig. 4b) or doxorubicin (Supplementary Fig. 11a) was drastically reduced, to the level of the *acrB*-deficient strain (Fig. 4b). Doxorubicin efflux activity was also completely lost in these loop translocation mutants (Fig. 4d).

We determined the crystal structure of one of these mutants (Gly616Pro Gly619Pro; Supplementary Fig. 13a, b) and found that the overall structure was not altered (Supplementary Fig. 13c). The Gly-to-Pro double mutations in the Phe-617 loop did not prevent the AcrB monomers from adopting different conformations. However, as shown in Supplementary Fig. 13d, the position of the mutant Phe-617 loop in the binding monomer is closer to that in the wild-type access monomer than that in the wild-type binding monomer, indicating that the movement of the Phe-617 loop during the conformational change from the access form to the binding form is prevented by the Gly616Pro Gly619Pro mutation as expected.

The CAVER protein-analysis software²⁹ estimates that there are three possible entrance channels: channel 1, from the vestibule¹⁵; channel 2, from the lower external depression^{23,25}; and channel 3, from the central cavity (Fig. 5a). Leu 674 is located at the juncture of channels 1 and 2, Ser 462 is halfway along channel 1, and Phe 666 and Ala 677 are located at the entrance of channel 2. Because the channels are narrow, the introduction of a high-volume side chain at these positions is expected to prevent drug entrance. The growth of strains bearing the mutation Leu674Trp, Ser462Lys, Phe666Trp or Ala677Trp was significantly decreased by subinhibitory concentrations of either erythromycin or doxorubicin (Fig. 4b and Supplementary Fig. 11a). The doxorubicin efflux activity of the Leu674Trp, Phe666Trp and Ala677Trp mutant strains was significantly reduced, and that of the Ser462Lys mutant strain was nearly undetectable (Fig. 4d). These results indicate that channels 1 and 2 act as a translocation route for HMMDs and LMMDs; however, we have not yet obtained site-directed mutagenesis data addressing the role of channel 3.

In conclusion, AcrB has two high-volume, multisite drug-binding pockets that line the drug translocation channel (Fig. 5a). Drugs taken up from the three possible entrances are transported by a peristaltic

mechanism³⁰ through both pockets and are finally extruded from the top exit (Fig. 5b). Whereas HMMDs are mainly recognized by the proximal pocket, the distal binding pocket recognizes LMMDs. This multiplicity of multisite drug-binding pockets and entrances contributes to the remarkably broad substrate specificity of AcrB.

METHODS SUMMARY

Crystallization and structure determination of drug-binding AcrB. C2 crystals of substrate-free AcrB and C2 crystals of AcrB–substrate complexes were generated (Methods). The structures of the substrate-free, rifampicin-bound, erythromycin-bound and rifampicin–minocycline-bound forms were solved at resolutions of 3.35, 3.35, 3.35 and 3.3 Å, respectively (Supplementary Table 1). The bound substrates were located using a difference Fourier map.

Site-directed mutagenesis and the measurement of the drug efflux activity of AcrB. Point mutations were introduced into the *acrB* gene harboured by the pBADacrB plasmid by using In-Fusion PCR cloning (Takara Bio Inc.) with mutagenic primers as described in Methods. The resulting mutant AcrB proteins have a polyhistidine tag on the carboxy terminus similarly to wild-type AcrB, and protein expression was measured using an anti-polyhistidine antibody. AcrB was expressed from a pBADacrB-derived plasmid in *acrB*-deficient MG1655 Δ *acrB* *E. coli* cells. AcrB expression was limited to roughly endogenous levels by adding the inducer arabinose (10 mM) during cell growth. The activity of the mutant AcrB proteins was measured using a drug susceptibility test for erythromycin (Methods) and a drug efflux assay. Drug efflux activity was measured by the quenching of doxorubicin fluorescence. When a 100- μ l *E. coli* cell suspension ($A_{600\text{ nm}}$ 18) was incubated with 28.6 μ M doxorubicin in a 96-well, flat-bottomed plate, the fluorescence intensity (excitation wavelength of 480 nm and emission wavelength of 600 nm) decreased in a time-dependent manner owing to the accumulation of doxorubicin in the cells. When doxorubicin was expelled from cells by the efflux pump, this decrease in fluorescence intensity was prevented. Therefore, the abrogation of doxorubicin fluorescence quenching is a good measure of efflux activity.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2010; accepted 14 October 2011.

Published online 27 November 2011.

- Ma, D., Cook, D. N., Hearst, J. E. & Nikaido, H. Efflux pumps and drug resistance in Gram-negative bacteria. *Trends Microbiol.* **2**, 489–493 (1994).
- Okusu, H., Ma, D. & Nikaido, H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J. Bacteriol.* **178**, 306–308 (1996).
- Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H. Jr. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301**, 75–100 (2000).

4. Nishino, K. & Yamaguchi, A. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.* **183**, 5803–5812 (2001).
5. Poole, K. Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *J. Mol. Microbiol. Biotechnol.* **3**, 255–264 (2001).
6. Nishino, K., Latifi, T. & Groisman, E. A. Virulence and drug resistance roles of multidrug efflux systems of *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.* **59**, 126–141 (2006).
7. Poole, K., Krebes, K., McNally, C. & Neshat, S. Multiple antibiotic resistance in *Pseudomonas aeruginosa*: evidence for involvement of an efflux operon. *J. Bacteriol.* **175**, 7363–7372 (1993).
8. Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **45**, 1126–1136 (2001).
9. Tikhonova, E. B. & Zgurskaya, H. I. AcrA, AcrB and TolC of *Escherichia coli* form a stable intermembrane multidrug efflux complex. *J. Biol. Chem.* **279**, 32116–32124 (2004).
10. Symmons, M. F., Bokma, E., Koronakis, E., Hughes, C. & Koronakis, V. The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proc. Natl Acad. Sci. USA* **106**, 7173–7178 (2009).
11. Fralick, J. A. Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of *Escherichia coli*. *J. Bacteriol.* **178**, 5803–5805 (1996).
12. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).
13. Zgurskaya, H. I. & Nikaido, H. AcrA is a highly asymmetric protein capable of spanning the periplasm. *J. Mol. Biol.* **285**, 409–420 (1999).
14. Mikolosko, J., Bobyk, K., Zgurskaya, H. I. & Ghosh, P. Conformational flexibility in the multidrug efflux system protein AcrA. *Structure* **14**, 577–587 (2006).
15. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443**, 173–179 (2006).
16. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).
17. Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H. & Koshland, D. E. Jr. Structural basis of multidrug-binding capacity of the AcrB multidrug efflux pump. *Science* **300**, 976–980 (2003).
18. Yu, E. W., Aires, J. R., McDermott, G. & Nikaido, H. A periplasmic drug-binding site of the AcrB multidrug efflux pump: a crystallographic and site-directed mutagenesis study. *J. Bacteriol.* **187**, 6804–6815 (2005).
19. Törnroth-Horsefield, S. *et al.* Crystal structure of AcrB in complex with a single transmembrane subunit reveals another twist. *Structure* **15**, 1663–1673 (2007).
20. Drew, D. *et al.* The structure of the efflux pump AcrB in complex with bile acid. *Mol. Membr. Biol.* **25**, 677–682 (2008).
21. Pos, K. M., Schiefner, A., Seeger, M. A. & Diederichs, K. Crystallographic analysis of AcrB. *FEBS Lett.* **564**, 333–339 (2004).
22. Murakami, S. Multidrug efflux transporter, AcrB—the pumping mechanism. *Curr. Opin. Struct. Biol.* **18**, 459–465 (2008).
23. Seeger, M. A. *et al.* Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **313**, 1295–1298 (2006).
24. Seeger, M. A. *et al.* Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. *Nature Struct. Mol. Biol.* **15**, 199–205 (2008).
25. Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grutter, M. G. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol.* **5**, e7 (2007).
26. Sennhauser, G., Bukowska, M. A., Briand, C. & Grutter, M. G. Crystal structure of the multidrug exporter MexB from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **389**, 134–145 (2009).
27. Schumacher, M. A. *et al.* Structural mechanisms of QacR induction and multidrug recognition. *Science* **294**, 2158–2163 (2001).
28. Bohnert, J. A. *et al.* Site-directed mutagenesis reveals putative substrate binding residues in the *Escherichia coli* RND efflux pump AcrB. *J. Bacteriol.* **190**, 8225–8229 (2008).
29. Medek, P., Benes, P. & Sochor, J. Computation of tunnels in protein molecules using Delaunay triangulation. *J. WSCG* **15**, 107–114 (2007).
30. Pos, K. M. Drug transport mechanism of the AcrB efflux pump. *Biochim. Biophys. Acta* **1794**, 782–793 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Tsukihara for advice on the crystallographic analysis. We also thank N. Kato for discussion about the organic chemistry of the drugs that were investigated in this study. Our diffraction data were collected using Osaka University's beamline BL44XU at SPring-8, which was equipped with an MX225-HE detector (Rayonix) and was financially supported by the Academia Sinica and the National Synchrotron Radiation Research Center (Taiwan). We are also grateful to the technical staff of the Comprehensive Analysis Center of the Institute of Scientific and Industrial Research for their assistance. This work was supported by the Program for the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation and Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions R.N. and K.S. performed the crystallographic analysis. S.Y. and K.N. performed the molecular biological and biochemical analyses. A.Y. designed the research and wrote the manuscript.

Author Information The coordinates for unliganded AcrB and the AcrB–rifampicin, AcrB–erythromycin and AcrB–rifampicin–minocycline complexes have been deposited in the Protein Data Bank under accession numbers 3AOA, 3AOB, 3AOC, and 3AOD, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.Y. (akihito@sanken.osaka-u.ac.jp).

METHODS

Preparation of native and drug-complexed crystals. Wild-type AcrB was tagged with polyhistidine at its C terminus as described previously¹⁶, and polyhistidine-tagged AcrB was prepared as described previously¹⁵. To ensure AcrB–drug complex formation, the drug-to-protein molar ratio was in the range of 10:1 to 30:1. Rifampicin (2.4 mM), erythromycin thiocyanate (7.1 mM) or minocycline (3.6 mM) plus rifampicin (4.7 mM) was added to the purified protein and incubated for several hours at 4 °C before set-up of the crystallization procedure. Crystals were grown using the vapour diffusion method from sitting drops at 25 °C. A protein solution containing 28 mg ml^{−1} protein, 20 mM sodium phosphate (pH 6.4), 10% (v/v) glycerol and 0.1% (w/v) sucrose monododecanoate was mixed with an equal volume of reservoir solution containing 14% (w/v) polyethylene glycol 4000, 100 mM sodium phosphate (pH 6.4) and 100 mM NaCl. Cryoprotection was achieved by raising the glycerol concentration stepwise to 30% (v/v) in 5% increments. Crystals were picked using LithoLoops (Protein Wave) for flash cooling in a cold nitrogen gas stream (100 K) from a cryostat (Oxford Cryosystems).

Crystallographic analysis. Each data set was collected using the BL44XU beamline at SPring-8 with an MX225-HE charge-coupled-device detector (Rayonix) at 100 K. The diffraction data were processed and scaled using the HKL2000³¹ package. The crystals belonged to the C2 space group, and the cell dimensions are listed in Supplementary Table 1. The initial phases were first determined by molecular replacement with MOLREP³² in the CCP4 program suite³³ using the atomic coordinates of AcrB (Protein Data Bank ID, 2DHH) as a search model. Model rebuilding was performed using COOT³⁴, and model refinement was performed using CNS³⁵ and REFMAC³⁶. Of the 1,049 residues in the entire molecule, the atomic parameters of residues 1–498 and 513–1,036 converged well during refinement. To identify drug molecules in the AcrB–drug complexed crystals, difference Fourier maps were calculated with coefficients of $(|F_{\text{drug}}| - |F_{\text{free}}|)\exp(i\alpha_{\text{free}})$ at a resolution of 3.35 Å, where F_{drug} is the structure factor in the presence of a bound substrate, F_{free} is the structure factor of the drug-free crystal and α_{free} is the phase of the drug-free crystal refined at a resolution of 3.35 Å. Refinement was performed at the same resolution for the rifampicin-bound complex and for the erythromycin-bound complex, and at a resolution of 3.30 Å for the minocycline–rifampicin-bound complex with constrained structures for the drug molecules. The quality of the diffraction data and the refinement statistics are provided in Supplementary Table 1. Each figure was constructed using PYMOL (version 1.3r1; Schrödinger, LLC).

Site-directed mutagenesis. The *acrB* gene was amplified from *E. coli* MG1655³⁷ genomic DNA using the Eco_acrB-F_SalI (5'-CGCGTCGACAGTCTTAAC TAAACAGGAGC-3') and Eco_acrB-R_SphI (5'-CGCGCATGCTCAATGGTG ATGGTGATGATGATCGACAGTATGGCT-3') primers, which introduce SalI and SphI restriction sites on each end of the amplified fragment (underlined in the primer sequences above). This PCR product was cloned into the SalI and SphI sites of the pBAD33 vector³⁸ to produce the pBADacrB plasmid. Point mutations were introduced using mutagenic primers to create the following codon replacements: S134A S135A (AGC→GCG AGC→GCG), S79R T91R (TCC→CGC ACC→CGC), D681W (GAC→TGG), F615A (TTC→GCG), F610A

(TTC→GCG), F628A (TTC→GCG), G614P G621P (GGC→CCG GGT→CCG), G616P G619P (GGC→CCG GGA→CCG), L674W (CTG→TGG), S462K (TCT→AAA), F666W (TTT→TGG) and A677W (GCA→TGG). The constructed plasmids were sequenced using a 3100-Avant Genetic Analyser (Applied Biosystems) to ensure the presence of the desired mutation. The resulting mutant AcrB proteins have a polyhistidine tag on the carboxy terminus similarly to wild-type AcrB, and protein expression was measured using an anti-polyhistidine antibody. AcrB was expressed from a pBADacrB-derived plasmid in *acrB*-deficient MG1655^{ΔacrB} *E. coli* cells. AcrB expression was limited to roughly endogenous levels by adding the inducer arabinose (10 mM) during cell growth.

Drug susceptibility tests for erythromycin. Single colonies of *acrB*-deficient *E. coli* strains that harboured pBADacrB-derived plasmids were inoculated into 2 ml of lysogeny broth. Cultures were grown overnight at 37 °C, and 2×10^7 c.f.u. μl^{−1} of bacteria was inoculated into 200 μl of lysogeny broth containing erythromycin (32 μg ml^{−1}) and 10 mM arabinose. Bacterial growth was measured using 96-well plates and a Versa Max microplate reader (Molecular Devices).

Drug efflux assay. Cultures of *E. coli* that were grown overnight were diluted in lysogeny broth to achieve a cell density of 10^7 cells ml^{−1}. The cells were grown under aerobic conditions at 37 °C in the presence of 10 mM arabinose as an inducer until $A_{600\text{ nm}}$ 0.3 was reached. The cells were then collected, washed twice with 100 mM potassium phosphate buffer (pH 7.5 and containing 5 mM MgSO₄) and then resuspended in the same buffer until $A_{600\text{ nm}}$ 18 was reached. Each assay was performed in 96-well, flat-bottomed black plates (IWAKI) using a final cell volume of 100 μl, and doxorubicin (Sigma-Aldrich) was added to a final concentration of 28.6 μM. For the competition assays, 28.6 μM minocycline, erythromycin, rifampicin or kanamycin was added. The fluorescence of doxorubicin was measured using an SH-8100 microplate reader (Corona Electric Co.) with an excitation wavelength of 480 nm and an emission wavelength of 600 nm. The fluorescence intensity decreased in a time-dependent manner owing to the accumulation of doxorubicin in the cells, and the decrease in fluorescence intensity was prevented when doxorubicin was expelled by an efflux pump.

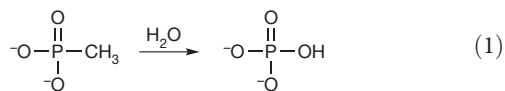
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

Intermediates in the transformation of phosphonates to phosphate by bacteria

Siddhesh S. Kamat¹, Howard J. Williams¹ & Frank M. Raushel¹

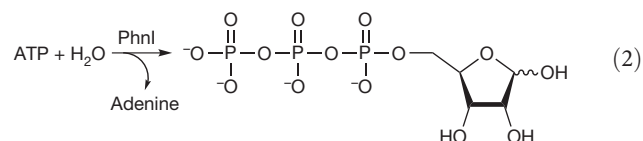
Phosphorus is an essential element for all known forms of life. In living systems, phosphorus is an integral component of nucleic acids, carbohydrates and phospholipids, where it is incorporated as a derivative of phosphate. However, most Gram-negative bacteria have the capability to use phosphonates as a nutritional source of phosphorus under conditions of phosphate starvation¹. In these organisms, methylphosphonate is converted to phosphate and methane. In a formal sense, this transformation is a hydrolytic cleavage of a carbon–phosphorus (C–P) bond, but a general enzymatic mechanism for the activation and conversion of alkylphosphonates to phosphate and an alkane has not been elucidated despite much effort for more than two decades. The actual mechanism for C–P bond cleavage is likely to be a radical-based transformation². In *Escherichia coli*, the catalytic machinery for the C–P lyase reaction has been localized to the *phn* gene cluster¹. This operon consists of the 14 genes *phnC*, *phnD*, ..., *phnP*. Genetic and biochemical experiments have demonstrated that the genes *phnG*, *phnH*, ..., *phnM* encode proteins that are essential for the conversion of phosphonates to phosphate and that the proteins encoded by the other genes in the operon have auxiliary functions^{1,3–6}. There are no functional annotations for any of the seven proteins considered essential for C–P bond cleavage. Here we show that methylphosphonate reacts with MgATP to form α -D-ribose-1-methylphosphonate-5-triphosphate (RPNTP) and adenine. The triphosphate moiety of RPNTP is hydrolysed to pyrophosphate and α -D-ribose-1-methylphosphonate-5-phosphate (PRPN). The C–P bond of PRPN is subsequently cleaved in a radical-based reaction producing α -D-ribose-1,2-cyclic-phosphate-5-phosphate and methane in the presence of S-adenosyl-L-methionine. Substantial quantities of phosphonates are produced worldwide for industrial processes, detergents, herbicides and pharmaceuticals^{7–9}. Our elucidation of the chemical steps for the biodegradation of alkylphosphonates shows how these compounds can be metabolized and recycled to phosphate.

The bacterial C–P lyase gene cluster enables alkylphosphonates to be converted to phosphate (equation (1)).

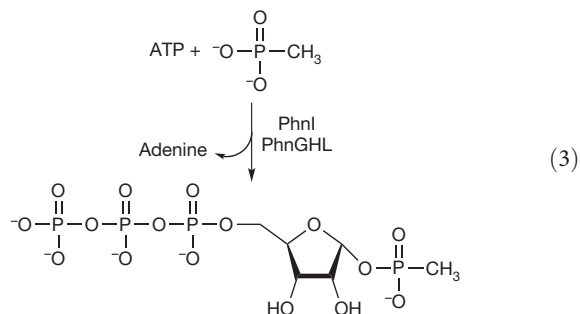


We cloned the genes *phnG*, *phnH*, ..., *phnM* and expressed them in *E. coli*, and purified to homogeneity the resulting proteins. PhnI, PhnJ, PhnK and PhnL were purified as amino-terminal glutathione S-transferase (GST) fusion proteins. PhnG, PhnH and PhnM were purified without a GST tag. Previous experiments have hinted that a ribose intermediate exists in the esterification of phosphonate substrates before conversion to phosphate^{10,11}. Because PhnI showed a distant relationship to enzymes that are functionally annotated as nucleosidases¹², we incubated this protein with a small, focused library of potential ribose donors. The liberation of the free base was followed spectrophotometrically at 240–350 nm in the presence of coupling enzymes that are capable of detecting the formation of adenine¹³,

guanine¹⁴, cytosine¹⁵ or xanthine¹⁶. The best substrates for PhnI are GTP ($k_{\text{cat}}/K_m = 8.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$; k_{cat} , turnover number; K_m , Michaelis constant) and ATP ($k_{\text{cat}}/K_m = 1.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$). The products are D-ribose-5-triphosphate (RTP) and the free base as shown for ATP in equation (2). The kinetic constants for the nucleosidase activity of PhnI with GTP and ATP are listed in Supplementary Table 1. The structure of RTP was identified by ³¹P NMR spectroscopy and further confirmed by multidimensional NMR.



The formation of RTP from either ATP or GTP did not seem to be a productive pathway for the transformation of phosphonates to phosphate. We therefore incubated phosphate and phosphonate derivatives with MgATP in the presence of PhnI to determine whether any of these compounds could displace adenine. RTP and adenine were the only products formed in the presence of either phosphate or phosphonate derivatives. There were no changes in the reaction products when we used the GST-tagged PhnI in the presence or absence of factor Xa¹⁷. Because it was previously postulated that the C–P lyase reaction may involve a multiprotein complex^{3,18}, we added additional proteins to the reaction mixture. PhnG, PhnH, PhnK and PhnL were incubated together with PhnI, MgATP and methylphosphonate, but no new products were detected. Because PhnI, PhnK and PhnL were all present as GST fusion proteins, factor Xa was added to the reaction mixture for *in situ* cleavage of the GST tags. After addition of factor Xa, a new resonance was observed by ³¹P NMR spectroscopy at a chemical shift of 17.8 p.p.m. (Fig. 1a). The NMR spectra are consistent with the formation of RPNTP, shown in equation (3).



To determine whether all of the proteins are required for this transformation, we removed each protein individually from the reaction mixture. The only protein that could be removed without affecting the transformation was PhnK. Thus, PhnI, in the presence of PhnG, PhnH and PhnL, is required to catalyse the nucleophilic attack of methylphosphonate on the anomeric carbon of MgATP to form adenine and RPNTP. The kinetic constants for the reaction of methylphosphonate with ATP ($k_{\text{cat}}/K_m = 3.5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) and GTP ($k_{\text{cat}}/K_m = 3.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) in the presence of PhnI, PhnG, PhnH and

¹Department of Chemistry, PO Box 30012, Texas A&M University, College Station, Texas 77843, USA.

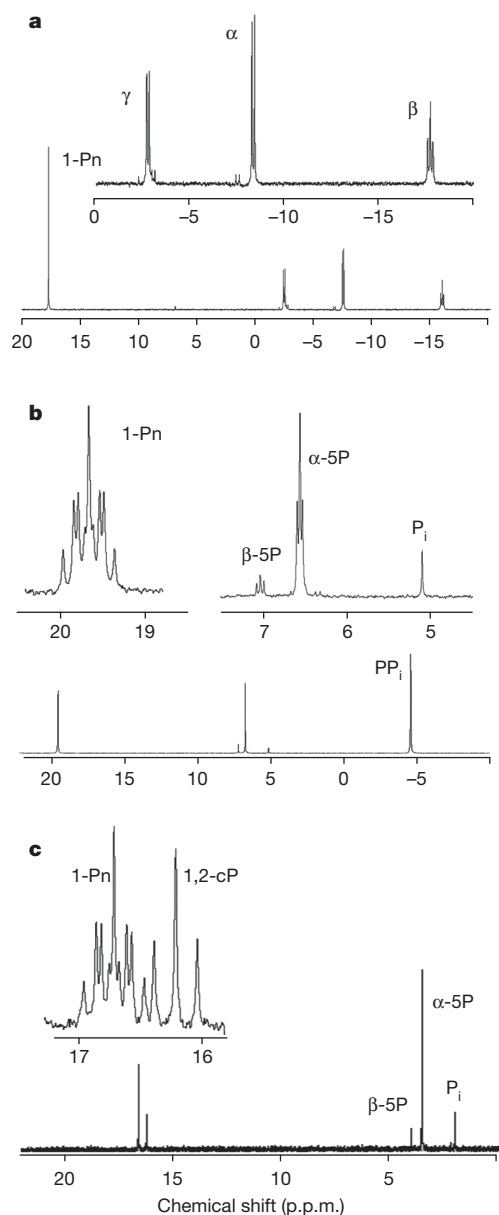
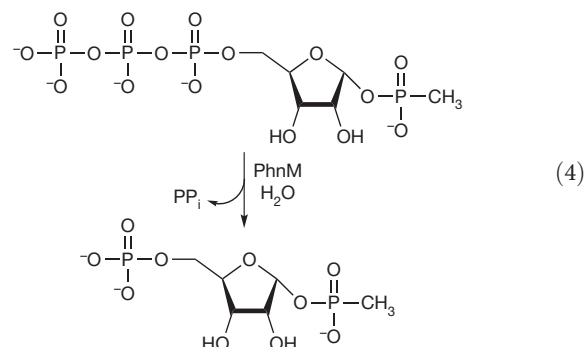


Figure 1 | ^{31}P NMR spectra of the reaction products catalysed by PhnI, PhnM and PhnJ. **a**, RPNTP from the reaction of methylphosphonate and MgATP catalysed by PhnI in the presence of PhnG, PhnH, and PhnL at pH 8.5. The methylphosphonyl group is labelled as 1-Pn. Inset, ^{31}P – ^{31}P coupling of the triphosphate portion of the RPNTP product (the α , β and γ phosphoryl groups). **b**, The formation of PRPn from RPNTP in the presence of PhnM. Inset, proton-coupled spectrum at pH 8.5 showing the multiplet that corresponds to 1-Pn and the triplet that corresponds to the 5-phosphate (5P). P_i , inorganic phosphate. **c**, The formation of PRcP from PRPn in the presence of PhnJ at pH 6.8. Inset, proton-coupled spectrum showing the formation of a new triplet that corresponds to the 1,2-cyclic moiety of PRcP (1,2-cP). The chemical shifts for the phosphate moiety at the fifth carbon atom of PRcP and PRPn are coincident with one another (3.4 p.p.m.).

PhnL are presented in Supplementary Table 1. We note that these proteins begin to precipitate after removal of the GST tags and that the kinetic constants are therefore not definitive. No reaction was observed with the fusion proteins, and the GST tags must be removed *in situ*. The stoichiometries of the four proteins required for complex formation and the individual functions of PhnG, PhnH and PhnL are uncertain.

Previous studies have suggested that PRPn is a potential intermediate in the conversion of alkylphosphonates to phosphate by *E. coli*^{6,10,11}.

Hence, it was rational to assume that one of the proteins expressed by the *phn* operon would catalyse the hydrolysis of the β - and γ -phosphoryl groups from RPNTP. PhnM is the prime candidate for this reaction because this protein is a member of the amidohydrolase superfamily, members of which are known to catalyse hydrolytic reactions at carbon and phosphorus centres¹⁹. We incubated PhnM with RPNTP in the presence of MgCl_2 and ZnCl_2 . The changes in the ^{31}P NMR spectrum are shown in Fig. 1b. The NMR spectra are consistent with the formation of PRPn and pyrophosphate as the major products. D-ribose-5-diphosphate and RTP are also substrates for PhnM. When the hydrolysis reaction of RTP catalysed by PhnM was conducted in ^{18}O -labelled water, the ^{18}O was found exclusively in D-ribose-5-phosphate and not in pyrophosphate. Therefore, water attacks the α -phosphoryl group of RTP rather than the β -phosphoryl group. The reaction catalysed by PhnM with RPNTP is presented in equation (4). The kinetic constants for the hydrolysis of RPNTP, RTP and D-ribose-5-diphosphate by PhnM are provided in Supplementary Table 1. RPNTP ($k_{\text{cat}}/K_m = 1.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) was the best substrate for PhnM and the kinetic constants did not change for the hydrolysis of RPNTP when PhnG, PhnH, PhnI, PhnK, PhnL and factor Xa were added to the assay mixture.



The deletion of *phnJ* from *E. coli* led to the identification of α -D-ribose-1-methylphosphonate in the growth medium, and this led to the prediction that PRPn is the ultimate substrate for the C–P lyase reaction⁶. This conjecture is consistent with our results because we can synthesize PRPn from ATP and methylphosphonate by the combined actions of PhnI, PhnG, PhnH, PhnL and PhnM. Of the remaining two enzymes expressed by the *phn* operon, the protein most likely to catalyse the C–P bond cleavage is PhnJ. This enzyme has four conserved cysteine residues, with a spacing of $\text{CX}_2\text{CX}_{21}\text{CX}_5\text{C}$ (where X_n denotes a sequence of some n amino acids), that could form part of an iron–sulphur (Fe–S) cluster¹². The cleavage of PRPn to α -D-ribose-1,2-cyclic-phosphate-5-phosphate (PRcP) is assumed to involve a radical mechanism and such reactions can be catalysed by radical SAM enzymes. These enzymes use a [4Fe–4S] cluster and S-adenosyl-L-methionine (SAM) to catalyse radical reactions and/or rearrangements²⁰. The radical SAM superfamily was originally thought to have a highly conserved $\text{CX}_3\text{CX}_2\text{C}$ motif as a signature for harbouring the [4Fe–4S] cluster²⁰, but with the successful reconstitutions of ThiC²¹, HmdA²² and Dph2²³ a more diverse combination of conserved cysteine residues can function for the assembly of the [4Fe–4S] cluster. If PhnJ incorporates a [4Fe–4S] cluster, it is not clear which of the four cysteine residues are required.

The as-purified GST–PhnJ fusion protein was blackish in colour. Inductively coupled plasma mass spectroscopy of this protein demonstrated the binding of 2.2 ± 0.2 equiv. of iron per PhnJ monomer. The aerobically purified protein was made anaerobic in a glove box by bubbling argon through the protein solution and then allowed it to equilibrate for 4 h. Reconstitution of the [4Fe–4S] cluster was initiated by the slow anaerobic addition of a 50-fold excess of FeSO_4 and Na_2S . The protein with the reconstituted Fe–S cluster had a reddish-brown colour. The ultraviolet–visible absorption spectrum of PhnJ,

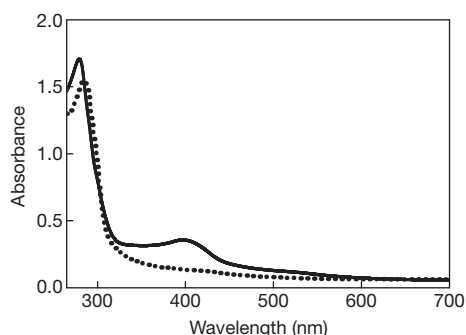
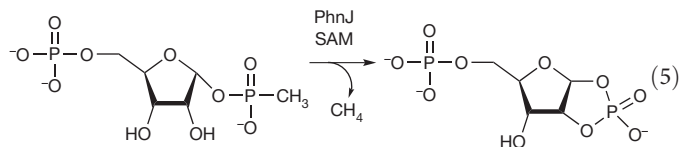


Figure 2 | Ultraviolet-visible absorbance spectrum of PhnJ (31 μ M) after anaerobic reconstitution of the Fe-S cluster (solid line). The peak at 280 nm is due to the protein and the absorbance centred at 403 nm represents the $[4\text{Fe-4S}]^{2+}$ cluster. The dotted line represents the absorbance spectrum of PhnJ (27 μ M) reconstituted with a $[4\text{Fe-4S}]^{2+}$ cluster after reduction of the cluster with sodium dithionite to the $[4\text{Fe-4S}]^+$ form.

reconstituted with iron and sulphide, had a broad absorption band centred at 403 nm that is indicative of a $[4\text{Fe-4S}]^{2+}$ cluster^{20–23} (Fig. 2). The absorption band disappears on addition of 1 mM dithionite (Fig. 2), suggesting the reduction of the $[4\text{Fe-4S}]^{2+}$ cluster to the $[4\text{Fe-4S}]^+$ species, which is the active form for most radical SAM enzymes^{20–23}.

There are two compounds that are potential substrates for cleavage by PhnJ of the C–P phosphonate bond: PRPn and RPnTP. Both of these compounds were incubated with 125 μ M PhnJ with a $[4\text{Fe-4S}]$ cluster, 2 mM SAM and 1 mM dithionite, anaerobically, at pH 6.8. We analysed the reactions by ^{31}P NMR spectroscopy but observed no change in the NMR spectrum for either substrate. When the reaction was supplemented with factor Xa, no change in the ^{31}P NMR spectrum with RPnTP was observed, but a new resonance appeared at 16.2 p.p.m. when PRPn was used as a substrate (Fig. 1c). The increase in the resonance at 16.2 p.p.m. correlated with a decrease in the phosphonate resonance of PRPn at 16.6 p.p.m. The new resonance splits into a triplet in the ^1H -coupled ^{31}P spectrum (Fig. 1c, inset), demonstrating that the product is no longer a methylphosphonate. The new resonance is consistent with a cyclic phosphate, and the proton coupling constant of 21 Hz indicates that the phosphate moiety of the product is esterified to the hydroxyl groups attached to the first

and second carbons atoms of the ribose. Thus, PhnJ requires a reduced $[4\text{Fe-4S}]$ cluster and SAM to catalyse the formation of PRcP from PRPn. The overall reaction is illustrated in equation (5). There was no reaction in the absence of added SAM.



We used gas chromatography and mass spectrometry to confirm the formation of methane. The $[4\text{Fe-4S}]$ -cluster-reconstituted PhnJ was incubated with 2 mM SAM, 1 mM dithionite, 5 mM PRPn and factor Xa in a sealed tube for 5 h. Gas chromatographic analysis of the headspace above the liquid showed a single peak that co-eluted with a methane standard (Supplementary Fig. 2). The formation of methane was confirmed by coupling the output of the gas chromatography to a mass spectrometer and detection of a molar mass of 16 g mol^{-1} (approximately that of methane). Thus, the two products that form from the action of PhnJ on PRPn are methane and PRcP. We were unable to detect the formation of methanol or formaldehyde in the reaction mixture.

The addition of SAM, dithionite, PhnJ and factor Xa were required for the formation of PRcP and methane from the $[4\text{Fe-4S}]^{2+}$ -cluster-reconstituted PhnJ. To determine the fate of SAM during the reaction, we used high-performance liquid chromatography and amino-acid analysis. Liquid chromatography showed the formation of 5'-deoxyadenosine, and amino-acid analysis confirmed the formation of methionine. These products were formed only when PhnJ (with a reconstituted Fe–S cluster), factor Xa, SAM, dithionite and PRPn were added to the reaction mixture. The omission of any one of the above components resulted in no formation of 5'-deoxyadenosine (Supplementary Fig. 3). The concentration of PRcP and methane formed in the presence of reconstituted PhnJ and SAM showed that one to four turnovers of product were formed per PhnJ monomer. The small number of turnovers per enzyme may reflect the poor solubility of PhnJ after proteolytic cleavage of the N-terminal GST fusion tag and the instability of the $[4\text{Fe-4S}]$ cluster. A working model for the conversion of PRPn to PRcP by PhnJ is presented in Fig. 3.

The reaction intermediates for the conversion of alkylphosphonates to phosphate in *E. coli* have now been identified. PhnI, in the presence

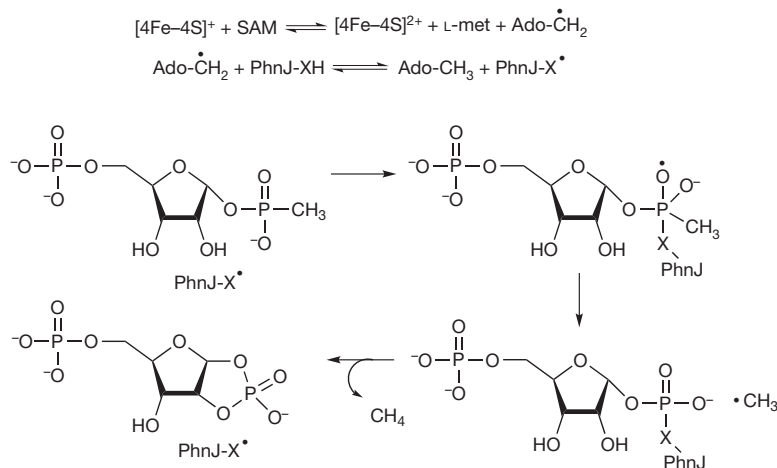


Figure 3 | Working model for the transformation of PRPn to PRcP. The cleavage of the C–P bond in PRPn by PhnJ reconstituted with a $[4\text{Fe-4S}]^{1+}$ cluster and SAM is probably initiated by electron transfer from the Fe–S cluster to reductively cleave SAM and thus transiently generate L-methionine (L-met) and a 5'-deoxyadenosyl radical (Ado-CH_2^\bullet). This radical may subsequently catalyse the formation of a protein radical (PhnJ-X^\bullet), presumably a cysteine-based thiyl radical. Thiyl radicals have previously been demonstrated in

pyruvate formate lyase^{24,25} and methyl coenzyme M reductase²⁶. The thiyl radical may attack the phosphonate moiety of the substrate to liberate a methyl radical with formation of a thioester intermediate. Intramolecular attack by the hydroxyl of the second carbon atom of the substrate would generate PRcP and the free thiol group. Methane would be formed through hydrogen atom abstraction from either 5'-deoxyadenosine or the putative cysteine residue.

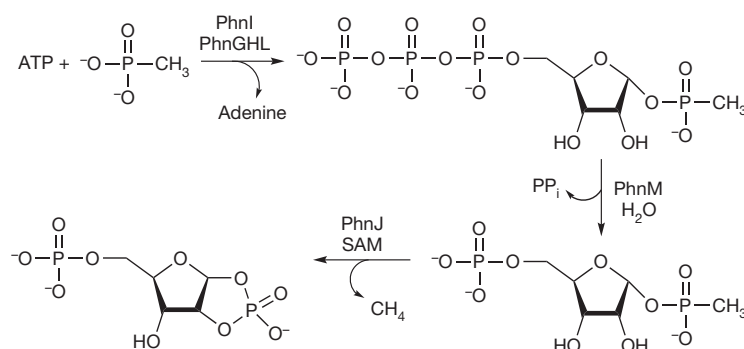


Figure 4 | Reaction pathway for the conversion of methylphosphonate to PRcP. The proteins PhnG, PhnH, PhnI, PhnJ, PhnL and PhnM are required for this transformation. The role of PhnK is unknown. PhnGHL denotes PhnG, PhnH and PhnL.

of PhnG, PhnH and PhnL, catalyses the formation of RPNTP from MgATP and methylphosphonate. PhnM catalyses the hydrolysis of RPNTP to pyrophosphate and PRPn. PhnJ can be reconstituted anaerobically with an Fe-S cluster using ferrous sulphate, sodium sulphide and sodium dithionite. The reconstituted PhnJ catalyses the SAM-dependent radical cleavage of the C-P bond of PRPn to form PRcP and methane. The transformations catalysed by the C-P lyase system in *E. coli* are summarized in Fig. 4.

METHODS SUMMARY

Protein expression and purification. All vectors bearing the appropriate genes were transformed in BL21 DE3 cells and inoculated into lysogeny broth medium. Cells were grown at 37 °C to an absorbance of $A_{600\text{ nm}}$ 0.4, after which the temperature was reduced to 18 °C. At $A_{600\text{ nm}}$ 0.6, the cells were induced with 0.5 mM isopropyl- β -D-thiogalactoside. Cells were grown for 18–20 h after induction. Overexpression was confirmed by SDS-polyacrylamide gel electrophoresis. Detailed descriptions of the protein purification protocols are presented in Supplementary Information. All GST-tagged proteins (PhnI, PhnJ, PhnK and PhnL) were purified on a GSTrap column (GE Healthcare; 5 ml) and all poly-His tag proteins (PhnG and PhnH) were purified on a HisTrap column (GE Healthcare; 5 ml) following the manufacturer's instructions. PhnM was purified using gel filtration with a High Load Superdex 200 26/60 prep grade column followed by anion exchange chromatography (ResourceQ; 6 ml). The chemical reconstitution of PhnJ is explained in detail in Supplementary Information.

Purification of reaction products and NMR spectroscopy. All reactions were filtered by centrifugation through a 10- or 30-kDa membrane (VWR Chemicals) to separate the enzymes from the reaction mixture. The filtrate (0.5 ml) was loaded onto a 1-ml ResourceQ column and eluted with 0.1–1 M ammonium bicarbonate over 20 ml using an AKTA Purifier 10 high-performance liquid chromatograph. The peak fractions were collected and vacuum dried at 4 °C. The products were analysed by ^1H , ^{13}C and ^{31}P NMR. The samples for ^{31}P NMR (D_2O , 85% H_3PO_4 corresponds to $\delta = 0.00$) were made in 50 mM HEPES, 10% D_2O , pH 8.5–8.8 (or pH 6.8 for PhnJ) and analysed with a Varian Unity Inova 500 MHz NMR spectrometer or a Bruker Avance III 400 MHz NMR spectrometer. ^1H and ^{13}C NMR spectra were acquired on a Bruker Avance III 500 MHz NMR spectrometer equipped with an HCN cryoprobe.

Received 2 May; accepted 10 October 2011.

Published online 16 November 2011.

1. Metcalf, W. W. & Wanner, B. L. Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*. *Gene* **129**, 27–32 (1993).
2. Ahn, Y., Ye, Q., Cho, H., Walsh, C. T. & Floss, H. G. Stereochemistry of carbon-phosphorus cleavage in ethylphosphonate catalyzed by C-P lyase from *Escherichia coli*. *J. Am. Chem. Soc.* **114**, 7953–7954 (1992).
3. Metcalf, W. W. & Wanner, B. L. Mutational analysis of an *Escherichia coli* fourteen-gene operon for phosphonate degradation, using TnphoA' elements. *J. Bacteriol.* **175**, 3430–3442 (1993).
4. Hove-Jensen, B., Rosenkrantz, T. J., Haldimann, A. & Wanner, B. L. *Escherichia coli* *phnN*, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. *J. Bacteriol.* **185**, 2793–2801 (2003).
5. Errey, J. C. & Blanchard, J. S. Functional annotation and kinetic characterization of PhnO from *Salmonella enterica*. *Biochemistry* **45**, 3033–3039 (2006).
6. Hove-Jensen, B., McSorley, F. R. & Zechel, D. L. Physiological role of *phnP*-specified phosphoribosyl cyclic phosphodiesterase in catabolism of organophosphoric acids by the carbon-phosphorus lyase pathway. *J. Am. Chem. Soc.* **133**, 3617–3624 (2011).
7. Ternan, N. G., McGrath, J. W., McMullan, G. & Quinn, J. P. Organophosphonates: occurrence, synthesis and biodegradation by microorganisms. *World J. Microbiol. Biotechnol.* **14**, 635–647 (1998).

8. Kononova, S. V. & Nesmeyanova, M. A. Phosphonates and their degradation by microorganisms. *Biochemistry (Mosc.)* **67**, 184–195 (2002).
9. White, A. K. & Metcalf, W. W. Microbial metabolism of reduced phosphorus compounds. *Annu. Rev. Microbiol.* **61**, 379–400 (2007).
10. Avila, L. Z., Draths, K. M. & Frost, J. W. Metabolites associated with organophosphonate C-P bond cleavage: chemical synthesis and microbial degradation of [^{32}P]-ethylphosphonic acid. *Bioorg. Med. Chem. Lett.* **1**, 51–54 (1991).
11. Frost, J. W., Loo, S., Cordeiro, M. L. & Li, D. Radical-based dephosphorylation and organophosphonate biodegradation. *J. Am. Chem. Soc.* **109**, 2166–2171 (1987).
12. Parker, G. F., Higgins, T. P., Hawkes, T. & Robson, R. L. *Rhizobium (Sinorhizobium) meliloti* *phn* genes: characterization and identification of their protein products. *J. Bacteriol.* **181**, 389–395 (1999).
13. Kamat, S. S. *et al.* Catalytic mechanism and three-dimensional structure of adenine deaminase. *Biochemistry* **50**, 1917–1927 (2011).
14. Maynes, J. T., Yuan, R. G. & Snyder, F. F. Identification, expression and characterization of the *Escherichia coli* guanine deaminase. *J. Bacteriol.* **182**, 4658–4660 (2000).
15. Hall, R. S. *et al.* Three dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry* **50**, 5077–5085 (2011).
16. Krenitsky, T. A., Neil, S. M., Elion, G. B. & Hitchings, G. H. A comparison of the specificities of xanthine oxidase and aldehyde oxidase. *Arch. Biochem. Biophys.* **150**, 585–599 (1972).
17. La Vallie, E. R., McCoy, J. M., Smith, D. B. & Riggs, P. Enzymatic and chemical cleavage of fusion proteins. *Curr. Protocols Mol. Biol.* Unit 16.4B (1994).
18. Jochimsen, B. *et al.* Five phosphonate operon gene products as components of a multi-enzyme complex of the carbon-phosphorus lyase pathway. *Proc. Natl Acad. Sci. USA* **108**, 11393–11398 (2011).
19. Seibert, C. M., & Raushel, F. M. Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* **44**, 6383–6391 (2005).
20. Frey, P. A., Hegeman, A. D. & Ruzicka, F. J. The radical SAM superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88 (2008).
21. Chatterjee, A. *et al.* Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily. *Nature Chem. Biol.* **4**, 758–765 (2008).
22. McGlynn, S. E. *et al.* Identification and characterization of a novel member of the radical AdoMet enzyme superfamily and implications for the biosynthesis of the Hmd hydrogenase active site cofactor. *J. Bacteriol.* **192**, 595–598 (2010).
23. Zhang, Y. *et al.* Diphthamide biosynthesis requires an organic radical generated by iron-sulphur enzyme. *Nature* **465**, 891–896 (2010).
24. Parast, C. V., Wong, K. K., Lewis, S. A. & Kozarich, J. W. Hydrogen exchange of the glycyl radical of pyruvate formate-lyase is catalyzed by cysteine 419. *Biochemistry* **34**, 2393–2399 (1995).
25. Buis, J. M. & Broderick, J. B. Pyruvate formate-lyase activating enzyme: elucidation of a novel mechanism for glycyl radical formation. *Arch. Biochem. Biophys.* **433**, 288–296 (2005).
26. Thauer, R. K. & Shima, S. Methane as fuel for anaerobic microorganisms. *Ann. NY Acad. Sci.* **1125**, 158–170 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Barondeau and his laboratory for use of the anaerobic chamber and for advice on the assembly of Fe-S clusters in radical SAM enzymes. We also thank C. Xu for help with some of the ^{31}P NMR spectra and S. Burrows for help with cloning *phnI* and *phnH*. This work was supported in part by the National Institutes of Health (GM93342, GM71790) and the Robert A. Welch Foundation (A-840). The cryoprobe for the NMR spectrometer was purchased with funds from the National Science Foundation (0840464).

Author Contributions S.S.K., H.J.W. and F.M.R. designed the experiments. S.S.K. and H.J.W. performed the experiments. S.S.K., H.J.W. and F.M.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.M.R. (raushel@tamu.edu).

CAREERS

COLUMN Postdoc committees can offer insight into industry careers **p.576**

ASK THE EXPERT Get incisive answers to career concerns go.nature.com/4oiv3k

NATUREJOBS For the latest career listings and advice www.naturejobs.com

E. ALCOCK/NYOP DIFFUSION



Jean Rossier (left), Barbara Demeneix (centre) and Moshe Yaniv (right) inspire passion and hard work in their lab members, students and postdocs.

AWARDS

Conscientious counsellors

Nature's mentoring awards honour three scientists in France.

BY DECLAN BUTLER

French research has traditionally suffered from a rigid hierarchical and sometimes patriarchal lab structure, in which powerful lab bosses have often hindered the autonomy of younger lab members. But some say that this is changing. Many leading scientists in France work tirelessly to help their younger colleagues to spread their own wings, and the public funding that is increasingly available to young scientists helps them to establish independent careers.

To celebrate progressive mentors and the examples they set, this year's *Nature* Awards for Mentoring in Science honour three researchers in France. The €10,000 (US\$13,200) lifetime-achievement award was shared between Moshe Yaniv, a molecular biologist and emeritus professor at the Pasteur Institute in Paris, and Jean Rossier, a neurobiologist at the City of Paris Industrial Physics and Chemistry Higher Education Institution. The €10,000 mid-career award went to Barbara Demeneix, a molecular developmental endocrinologist at the National Museum of Natural History in Paris. Winners received the awards at a ceremony at the British Council in Paris on 12 December. (The author of this article was one of the five awards judges.)

As in previous years, candidates were nominated by their protégés; past students and postdocs emphasized how their mentors shaped their careers and nurtured their scientific skills. Fatima Mechta-Grigoriou, who now runs a stress and cancer laboratory at the Curie Institute in Paris, says of her former mentor, Yaniv: "He deeply influenced me, without me being conscious of this at that time, the way I currently drive my research laboratory and my own students and postdocs."

The mentors were lauded for their contagious passion and enthusiasm for science, and their hard work, rigour and fairness in sharing credit. Many protégés have sought to reproduce their mentors' behaviour, creating labs that are scientifically challenging and stimulating but also humane and nurturing — and giving their lab members enough autonomy to prepare them to run their own labs.

WORK TO LIVE

"For me," says Rossier, "the most important quality of a mentor is to be enthusiastic about the proposals and ideas of young investigators. Critiques should come later. First you should say 'do it' and then see if the new observation is real, reproducible and of interest." He strives to foster research projects aimed at "breakthroughs",

rather than focusing on incremental developments within established concepts.

For the award winners, doing excellent science need not come at the expense of having a personal life. All three are demanding bosses, and lab hours are often long, but they eschew the notion of a 24/7 lab (see *Nature* 477, 20–22; 2011 and J. Overbaugh *Nature* 477, 27–28; 2011). Catherine Chanfreau-Coffinier, a cardiologist researcher at the University of California, Los Angeles (UCLA), and a former graduate student with Yaniv, marvelled at how he and his wife, also a scientist, found time for family. "It has been very inspiring for me as a young scientist and a mother of two," she says. Samantha Richardson, a biochemist at the RMIT University in Melbourne, Australia, and a former researcher in Demeneix's lab, tells how her mentor, who has two children, advised her on how to juggle her career with "the conflicting and competing time constraints of a research career and small children, which can be stressful and often exasperating". Demeneix, the first-ever female chaired professor at the Paris museum, also prides herself on demonstrating how to "carry out research while accepting senior administrative responsibilities".

The winners often paid particular attention to foreign students' difficulties, from visa issues ►

► to lodgings; and they recognized protégés' needs. Yaniv can be tough but he "never forgets that the people doing the experiments are people, with human sensitivities and concerns", says Jonathan Weitzman, an epigenetics researcher at Paris Diderot University and a former postdoc with Yaniv.

The mentoring did not end when protégés left the lab. Carlos Cepeda, a UCLA neurophysiologist and a former postdoc with Rossier, says Rossier still influences him after 30 years, offering "scientific or even personal advice" whenever he reaches out. Paulina Damdimopoulou, now an endocrinologist at the Karolinska Institute in Stockholm, says that Demeneix continues to mentor her and "gives [her] opinions and advice without force-feeding solutions".

RESEARCH FREEDOM

Many former students say that their mentors gave them extreme freedom, encouraged ownership of projects and avoided micromanagement — while still understanding every detail of their work. "I tried to step back in order to let them interpret their own data, come up with conclusions and put forward further hypotheses," says Yaniv. "I endorsed their own initiatives and even let them embark in a wrong direction for a short while just to teach them how to be critical and analytical." Demeneix pays attention to major stumbling blocks such as unexpected experimental results, or times when nothing seems to work. "It is at these moments," she says, "that one needs to show maximum confidence in the student, teach them how to 'troubleshoot' experiments ... and to identify points in experiments where errors can occur."

The winning mentors all had multidisciplinary labs, in which managing levels of freedom among team members can be a particular challenge. Rossier notes that because no one neuroscientist can master all the field's techniques, lab members with different backgrounds have to work together. The challenge for the lab head, he says, is to "find the right balance between the singularity of each individual and the common goal of the group". Yaniv says that he tried to "create a lab space and atmosphere that increased the interaction between lab members".

To help graduate students and postdocs to acquire mentoring skills themselves, Yaniv gets them to co-supervise and mentor master's or first-year PhD students, and interns. "One becomes much more responsible and socially interactive, less selfish and more tolerant to others' scientific weaknesses," says Weitzman, "when one has some responsibilities toward others." ■

Declan Butler is a senior reporter at *Nature*, based in France.

COLUMN

Testing the waters

Postdoc committees can give insight into industry career paths, argue **Christopher Tsang** and **Michael Fisher**.

Graduates are often faced with a dilemma: stay in academia or seek a job in industry? One of us (C. T.) met this conundrum after he earned his biochemistry PhD in 2007. With little knowledge of the biotechnology industry, he opted for more training, taking a postdoc position at the University of California, Berkeley. If the postdoc did not work out, he thought, he could investigate companies. Even so, he had difficulty learning what working in industry would be like.

Many graduates and postdocs know surprisingly little about what an industry job entails. So we developed a mechanism that helps postdocs at Berkeley — and could help others.

It is called the Postdoc Industry Exploration Program (PIEP; see *Nature* **478**, 277; 2011), and is run by the Berkeley Postdoctoral Association. Influenced by an industry-exploration programme that enrolled 37 science postdocs at Massachusetts General Hospital in Boston (see *Nature Biotechnol.* **28**, 625–626; 2010), PIEP was first launched in January 2011. It organizes visits to biotechnology companies to showcase their organization, research and work atmospheres, and allows postdocs to network with employees and recruiters to gain useful contacts and a feel for company culture. Postdocs are inspired by employees who have made the transition from academia, and get to see team-oriented environments and cross-departmental collaborations. Companies gain access to a highly qualified pool of potential recruits.

PIEP workshops train postdocs in creating a professional image and conducting informative interviews with contacts — particularly useful for international postdocs unfamiliar with US business customs. Postdocs also learn how to market themselves, on paper and in person. Seminars provide information about career opportunities and US work-visa requirements.

The PIEP committee surveyed 55 postdocs who participated in site visits in the first year of the scheme. Only 42% were very interested in industry before the visits, but that rose to 73% afterwards. This success led Berkeley's vice-chancellor for research, who had funded PIEP's pilot year, to commit another year of support, including funding for a programme manager.

Postdocs who want to start a similar programme need only a few motivated colleagues and contacts at companies and within the university. The Berkeley PIEP committee is made up of four postdocs, who each volunteer



OCEAN/CORBIS

two hours of their time per week; each committee member organizes a site visit with a company's human-resources department and staff scientists. We have had to work closely together, offer support, exchange ideas and share responsibilities, while accommodating each other's heavy research schedules.

Some universities may not have the infrastructure to support or sustain industry-exploration programmes, but this is not an insurmountable obstacle. Postdocs can create such programmes by partnering with schemes that already receive university support. For example, Berkeley's department of molecular and cell biology sponsors a course that brings in doctoral-degree holders to discuss how they moved into a non-academic career. Although this course is aimed at graduate students, it is available to postdocs, and could be part of an industry-exploration programme. Postdocs at neighbouring institutions could also coordinate and pool resources to cover expenses for site visits. Instead of hiring a full-time programme manager, universities could offer stipends to postdocs willing to help out. Interested postdocs will probably still have to donate their time until their university provides financial support.

Postdocs want to make an informed choice between career paths. In our view, the keys to helping them are motivation and institutional support. Initiatives such as PIEP can benefit career-conscious postdocs struggling to sort out their next move in a tough economy. ■

Christopher Tsang and **Michael Fisher** are postdocs at the University of California, Berkeley, and co-founders of the Postdoc Industry Exploration Program.

► to lodgings; and they recognized protégés' needs. Yaniv can be tough but he "never forgets that the people doing the experiments are people, with human sensitivities and concerns", says Jonathan Weitzman, an epigenetics researcher at Paris Diderot University and a former postdoc with Yaniv.

The mentoring did not end when protégés left the lab. Carlos Cepeda, a UCLA neurophysiologist and a former postdoc with Rossier, says Rossier still influences him after 30 years, offering "scientific or even personal advice" whenever he reaches out. Paulina Damdimopoulou, now an endocrinologist at the Karolinska Institute in Stockholm, says that Demeneix continues to mentor her and "gives [her] opinions and advice without force-feeding solutions".

RESEARCH FREEDOM

Many former students say that their mentors gave them extreme freedom, encouraged ownership of projects and avoided micromanagement — while still understanding every detail of their work. "I tried to step back in order to let them interpret their own data, come up with conclusions and put forward further hypotheses," says Yaniv. "I endorsed their own initiatives and even let them embark in a wrong direction for a short while just to teach them how to be critical and analytical." Demeneix pays attention to major stumbling blocks such as unexpected experimental results, or times when nothing seems to work. "It is at these moments," she says, "that one needs to show maximum confidence in the student, teach them how to 'troubleshoot' experiments ... and to identify points in experiments where errors can occur."

The winning mentors all had multidisciplinary labs, in which managing levels of freedom among team members can be a particular challenge. Rossier notes that because no one neuroscientist can master all the field's techniques, lab members with different backgrounds have to work together. The challenge for the lab head, he says, is to "find the right balance between the singularity of each individual and the common goal of the group". Yaniv says that he tried to "create a lab space and atmosphere that increased the interaction between lab members".

To help graduate students and postdocs to acquire mentoring skills themselves, Yaniv gets them to co-supervise and mentor master's or first-year PhD students, and interns. "One becomes much more responsible and socially interactive, less selfish and more tolerant to others' scientific weaknesses," says Weitzman, "when one has some responsibilities toward others." ■

Declan Butler is a senior reporter at *Nature*, based in France.

COLUMN

Testing the waters

Postdoc committees can give insight into industry career paths, argue **Christopher Tsang** and **Michael Fisher**.

Graduates are often faced with a dilemma: stay in academia or seek a job in industry? One of us (C. T.) met this conundrum after he earned his biochemistry PhD in 2007. With little knowledge of the biotechnology industry, he opted for more training, taking a postdoc position at the University of California, Berkeley. If the postdoc did not work out, he thought, he could investigate companies. Even so, he had difficulty learning what working in industry would be like.

Many graduates and postdocs know surprisingly little about what an industry job entails. So we developed a mechanism that helps postdocs at Berkeley — and could help others.

It is called the Postdoc Industry Exploration Program (PIEP; see *Nature* **478**, 277; 2011), and is run by the Berkeley Postdoctoral Association. Influenced by an industry-exploration programme that enrolled 37 science postdocs at Massachusetts General Hospital in Boston (see *Nature Biotechnol.* **28**, 625–626; 2010), PIEP was first launched in January 2011. It organizes visits to biotechnology companies to showcase their organization, research and work atmospheres, and allows postdocs to network with employees and recruiters to gain useful contacts and a feel for company culture. Postdocs are inspired by employees who have made the transition from academia, and get to see team-oriented environments and cross-departmental collaborations. Companies gain access to a highly qualified pool of potential recruits.

PIEP workshops train postdocs in creating a professional image and conducting informative interviews with contacts — particularly useful for international postdocs unfamiliar with US business customs. Postdocs also learn how to market themselves, on paper and in person. Seminars provide information about career opportunities and US work-visa requirements.

The PIEP committee surveyed 55 postdocs who participated in site visits in the first year of the scheme. Only 42% were very interested in industry before the visits, but that rose to 73% afterwards. This success led Berkeley's vice-chancellor for research, who had funded PIEP's pilot year, to commit another year of support, including funding for a programme manager.

Postdocs who want to start a similar programme need only a few motivated colleagues and contacts at companies and within the university. The Berkeley PIEP committee is made up of four postdocs, who each volunteer



OCEAN/CORBIS

two hours of their time per week; each committee member organizes a site visit with a company's human-resources department and staff scientists. We have had to work closely together, offer support, exchange ideas and share responsibilities, while accommodating each other's heavy research schedules.

Some universities may not have the infrastructure to support or sustain industry-exploration programmes, but this is not an insurmountable obstacle. Postdocs can create such programmes by partnering with schemes that already receive university support. For example, Berkeley's department of molecular and cell biology sponsors a course that brings in doctoral-degree holders to discuss how they moved into a non-academic career. Although this course is aimed at graduate students, it is available to postdocs, and could be part of an industry-exploration programme. Postdocs at neighbouring institutions could also coordinate and pool resources to cover expenses for site visits. Instead of hiring a full-time programme manager, universities could offer stipends to postdocs willing to help out. Interested postdocs will probably still have to donate their time until their university provides financial support.

Postdocs want to make an informed choice between career paths. In our view, the keys to helping them are motivation and institutional support. Initiatives such as PIEP can benefit career-conscious postdocs struggling to sort out their next move in a tough economy. ■

Christopher Tsang and **Michael Fisher** are postdocs at the University of California, Berkeley, and co-founders of the Postdoc Industry Exploration Program.

natureOUTLOOK

TRADITIONAL ASIAN MEDICINE

22/29 December 2011 / Vol 480 / Issue No. 7378



COVER ART: NIK SPENCER

Editorial

Herb Brody,
Michelle Grayson,
Tony Scully

Art & Design

Wes Fernandes,
Alisdair MacDonald,
Andrea Duffy

Production

Karl Smart, Emilia
Orviss, Leonora
Dawson-Bowling

Sponsorship

Yvette Smith,
Yuki Fujiwara,
Gerard Preston

Marketing

Elena Woodstock,
Hannah Phipps

Project Manager

Helen Anthony

Art Director

Kelly Buckheit Krause

Magazine Editor

Tim Appenzeller

Editor-in-Chief

Philip Campbell

Editorial advisor

Felix Cheung

When the topic of traditional Asian medicine was first mooted, we were sceptical. To a magazine based in Europe and steeped in the history of science, there is much about traditional Asian medical practice that seems mystical and pseudoscientific. Other than well known success stories — artemisinin for malaria, and arsenic trioxide for leukaemia — there seemed to be a lack of scientifically proven remedies.

Yet a bit of probing revealed what a complex story this is. Not only are big efforts underway to modernize traditional medicine in China and Japan, but Western medicine is adopting some aspects of the Eastern point of view too. In particular, modern medical practitioners are coming around to the idea that certain illnesses cannot be reduced to one isolatable, treatable cause. Rather, a fall from good health often involves many small, subtle effects that create a system-wide imbalance.

But do traditional medicines actually work? Their personalized nature makes randomized controlled trials — the gold standard for testing drugs — extremely difficult. Rarely are two formulations identical. However, as modern medicine becomes more personalized, using biological and genetic markers, it is inadvertently developing the tools to better test traditional medicines.

Although artemisinin and arsenic trioxide are the archetypal examples of successful modern medicines mined from traditional Asian medicine, they do not represent the ideal convergence of the two systems. There are unique aspects to traditional Asian medicine that could hold great promise if they are artfully investigated. The goal of science should be to rigorously test each claim and sort the medical wheat from the pseudoscientific chaff.

We acknowledge the financial support of Saishunkan Pharmaceutical Co., Ltd and Kitasato University Oriental Medicine Research Center in producing this *Outlook*. As always, *Nature* takes full responsibility for all editorial content.

Michelle Grayson

Associate Editor, *Nature Outlooks*.

CONTENTS

S82 TCM

Made in China

A persisting practice through the times

S84 CONVERGENCE

Where West meets East

Common objectives, but different ways

S87 PERSPECTIVE

All systems go

Jan van der Greef

S88 MICROBIOME

That healthy gut feeling

When medicine is your cup of tea

S90 MODERNIZATION

One step at a time

Applying scientific standards to TCM

S93 PATENTS

Protecting China's national treasure

Building a wall around property

S94 MODERN TCM

Enter the clinic

What's in store for today's patient?

S96 JAPAN

Will the sun set on Kampo?

A shadow is cast across the country

S97 PERSPECTIVE

Herbal danger

Masatomo Sakurai

S98 REGULATIONS

Herbal medicine rule book

Turning a new page in the West

S100 PERSPECTIVE

The clinical trial barriers

Liang Liu *et al.*

S101 BIODIVERSITY

Endangered and in demand

What's driving demand?

COLLECTION

S105 The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine
Youyou Tu

S109 Adenosine A1 receptors mediate local anti-nociceptive effects of acupuncture
Nanna Goldman et al.

S115 Steroid-like compounds in Chinese medicines promote blood circulation via inhibition of Na⁺/K⁺-ATPase
Ronald J.Y. Chen et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl, Sxx–Sxx (2011). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Traditional Asian Medicine* supplement can be found at http://www.nature.com/nature/outlook/asian_medicine/.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group – Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2011 Nature Publishing Group

TCM

Made in China

Although modern medicine is established in Asia, traditional medicine also plays a big role in people's healthcare — and is gaining in popularity in other countries too.

BY FELIX CHEUNG

Traditional medicine — a system of ancient medical practice that differs in substance, methodology and philosophy to modern medicine — plays an important role in health maintenance for the peoples of Asia, and is becoming more frequently used in countries in the West. Despite their growing popularity, there are misunderstandings about what these traditional medicines comprise and the standards they conform to. Here we aim to clear up some of the common misconceptions.

Within Asia, traditional Chinese medicine (TCM) is the system with the longest history. TCM was developed through thousands of years of empirical testing and refinement. It was the only medical practice in China before the early nineteenth century, when English missionaries arrived, bringing with them the drugs, devices and practices of modern medicine.

Outside China, in other Asian countries including Japan, South Korea, Malaysia and Vietnam, traditional medicine has formed its own distinct culture. In Japan, the traditional

medicine is commonly called *Kampo*. In these countries, different traditional medicines might use different prescriptions or methods of diagnosis, but the underlying philosophy and principles are similar because they all originate in China.

TCM encompasses a wide range of practices, including some that are familiar to the West, such as herbal medicine and acupuncture, plus others that remain peculiar to most Westerners, such as cupping (heated cup therapy), *tuina* (massage), *qigong* (movement and breathing exercises) and *moxibustion* (burnt mugwort therapy). Investigating whether these therapies have underlying mechanisms of action is now a central task in TCM research. This *Nature Outlook* will focus mainly on herbal medicines, which are the most comparable to modern pharmaceuticals.

TCM TREATMENT

A visit to a TCM practitioner involves four examinations. First, the patient's skin complexion, physique and tongue condition are inspected. The TCM doctor then listens to the patient's voice to see if there are any breathing

problems, cough or phlegm, and sniffs to detect any body odours that might indicate ill health.

Next, the practitioner questions how the patient feels overall: if they are hot or cold; whether they are sweating; how their stools look; if the patient is thirsty, and so forth. Finally, the TCM practitioner palpates the patient's wrist to feel the quality of the pulse and so assess overall health.

It is the pattern of all these signs and symptoms that determines the diagnosis, and the practitioner would then prescribe a treatment regimen consisting of any or all of the above-mentioned therapies to resolve the problem. Within TCM, the healthy human body is viewed as an entity in equilibrium — the ultimate goal of treatment is to restore the *qi* (energy) and *yin-yang* (balance) of this complex system. Although the concept might seem strange to Western perceptions, many TCM practitioners draw a parallel with the well-understood scientific concepts of metabolism (roughly equivalent to energy), and immunity and homeostasis (balance) (see 'All systems go', page S87).

Most people who visit a TCM practitioner

ANCIENT CHINA

The system of traditional Chinese medicine has developed over thousands of years and features many classic texts that are still read today.

Huangdi's Internal Classic, one of the earliest collections of TCM, was probably completed during the Warring States period. It is still regarded as a must-read for all TCM students and practitioners.

The Yellow Emperor, Huangdi, is shown presenting books containing medical knowledge to Lei Gong, the Thunder Duke (right).



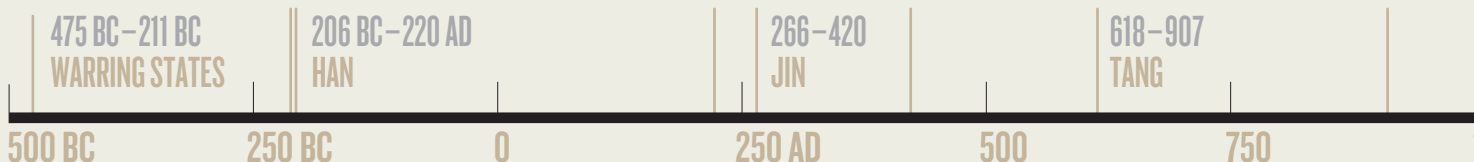
Two important books were written in the Han dynasty: Shennong's *Classic of Materia Medica*, — the first book to cover medicinal herbs, minerals and animal parts in the first century AD, and Zhang Zhongjing's *Treatise on Cold Damage and Miscellaneous Diseases* in 200 AD.



Clockwise from top left: Wang Shuhe and an excerpt from his *Pulse Classic*; Shennong Yandi (Red Emperor); Huangfu Mi.



In the Jin dynasty, Huangfu Mi compiled *The A-B Classic of Acupuncture and Moxibustion* (259 AD) and Wang Shuhe compiled the *Pulse Classic* (300 AD).



do so because they are feeling unwell. However, others visit a TCM practitioner while perfectly healthy. The reason is that TCM stresses the importance of preventive treatment — that is, its practitioners claim to determine a patient's state of *qi* and, if it is thought to be weak, attempt to strengthen it before a problem surfaces.

TCM PREVALENCE

Millions of patients around the world use TCM or a related practice. In Hong Kong and mainland China, approximately 60% of the population has consulted traditional medicine practitioners at least once. According to the latest national survey data, anywhere from 60% to 75% of the populations of Taiwan, Japan, South Korea and Singapore use traditional medicine at least once a year.

Elsewhere around the world, the use of TCM is also common. The United States is the biggest importer of TCM products from China, having spent US\$7.6 billion in 2010. According to the Centers for Disease Control and Prevention in Atlanta, Georgia, four out of ten US adults use complementary and alternative medicine (CAM) in a given year. The CAM category includes TCM and other herbal medicines, as well as non-traditional medicines such as homeopathy — so using CAM as a proxy for TCM might overestimate its use, but it does indicate a general willingness to seek alternative forms of therapy.

Europe, too, is experiencing a growth in the use of TCM. In 2010, exports of TCM products from China to European countries amounted to nearly US\$2 billion, and that figure is rising at 10% per year. National surveys have found

that, in the United Kingdom, approximately one in ten people had used CAM in the past year, whereas in Switzerland and the Scandinavian countries, the equivalent figures were between three and five out of ten.

Australia has even higher figures: a national survey found that two-thirds of the population had used CAM in the past year and, at 69 million, the number of visits by Australian adults to CAM practitioners was almost identical to the number of visits made to medical doctors.

PROSPECTS OF TCM RESEARCH

Modern medicine often clashes with traditional medicine such as TCM because of the inherent difficulties both sides have of appreciating the others' principles and concepts. In an effort to promote integration of traditional medicine

Investigating whether these therapies have underlying mechanisms of action is now a central task in TCM research.

with modern medicine, the World Health Organization (WHO) endorsed an international agreement drawn up in Beijing in November 2008 to support the safe and effective use of traditional medicine within the modern healthcare systems of member states.

Countries are pursuing this type of integration in various ways. For instance, between 2006 and 2010, Malaysia opened 12 hospitals that practice both modern and traditional medicine. In addition, the WHO has established 25

collaborating centres for traditional medicine: 7 in China, 5 in Africa, 3 in Europe, 2 in each of Japan, South Korea, India and the United States, and 1 in both North Korea and Vietnam. These centres aim to encourage research into traditional medicine, provide professional advice to support the development of WHO guidelines and, if necessary, provide training.

Pharmaceutical companies are also taking an active interest in TCM research. For instance, London-based GlaxoSmithKline has established a research and development base in Shanghai, China, and is actively seeking to expand its operations in traditional medicine. Most of these companies are hoping to find the next 'miracle' drug like artemisinin, an antimalarial drug extracted from the medicinal plant sweet wormwood, which has saved millions of lives.

Ensuring the safety, efficacy and quality of TCM requires a great deal of research and development. Investment from the central Chinese government is substantial. The total funding allocated to TCM research was 4.9 billion yuan (US\$770.5 million) in 2010 — more than quadruple its 2001 level. In 2010, according to the State Administration of Traditional Chinese Medicine, 6,093 Chinese scientists were employed in TCM research — a 53% increase since 2001.

As of May 2011, China has signed 91 TCM partnership agreements with more than 70 countries, which aim to promote greater recognition of TCM around the world. ■

Felix Cheung is editor of *Nature China* in Hong Kong.

During the peaceful Tang Dynasty, TCM spread to other countries and continued to evolve. In 659, the Chinese government issued the *Newly Revised Materia Medica*, officially replacing Shennong's *Classic*.

In 1578, Li Shizhen compiled *The Compendium of Materia Medica*, the most comprehensive documentation of the use of medicinal herbs, minerals and animal parts.



Protestant missionaries (below) arrive in 1807, bringing Western doctors and medical knowledge. The rise of Western medicine accelerated after the Opium Wars in 1839–1842 and 1856–1860 (left).

The Qing dynasty collapsed in 1911 and the Republic of China was founded. Foreign powers had a big influence in China, and favoured a Western school of thought over TCM.

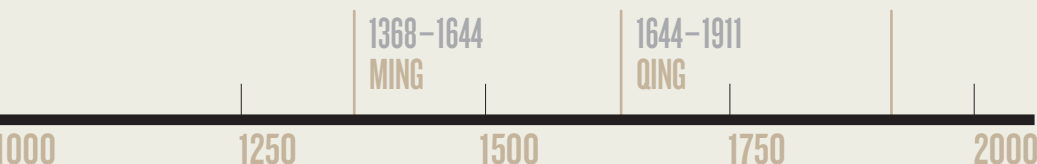
MODERN CHINA

In 1967, Mao Zedong initiated Project 523 that led to the development of artemisinin to combat malaria, which reinvigorated TCM development. In 2003, when severe acute respiratory syndrome (SARS) hit China, TCM was used to fend off SARS, giving further encouragement.

Below: A patient receives moxibustion in modern-day China.



Two pages from Li Shizhen's TCM encyclopaedia.





CONVERGENCE

Where West meets East

The concepts of Asia's traditional medicines might sound alien to Western ears, but some of them are starting to evolve to fit scientific investigation.

BY PENG TIAN

For around 200 years, two very different systems of medicine have been used in Asia to cure diseases and keep people healthy. The local Asian one is based on traditional Chinese medicine (TCM) — herbal mixtures developed through observation and experience accumulated over thousands of years, but with unknown mechanisms of action. On the other hand, modern medicine, imported from the West, consists of chemically purified compounds that have been discovered through scientific investigation and tested in controlled clinical trials. They differ in the composition of their medicines, the process of diagnosis, methods of proving a treatment's effectiveness, and even in their concept of 'health' (see 'Made in China', page S82).

Are these differences irreconcilable? Or, if TCM is modernized to the point where it can make scientifically valid claims, might it offer new perspectives that can benefit modern medicine — including clues on how to tackle the least tractable diseases and conditions? And likewise, can new perspectives being advanced in the West, such as systems biology, help lay a scientific foundation for TCM?

SPIRITUALITY IN MEDICINE

Although modern medicine has its roots in herbalism and in ancient Greek traditions that share many similarities with TCM, the practice of medicine was transformed by the Enlightenment and the consequent revolution in science and technology. Since the late eighteenth century, Western-style medicine has incorporated knowledge of anatomy, physiology, chemistry and biology, and its methods are evidence-based. TCM, although it is starting to take on these attributes, still relies heavily on ancient records and traditional practices.

TCM includes many tenets derived from Taoism, Confucianism and ancient Indian philosophies that describe the natural world, life and the human body. Concepts include yin and yang, which represent opposing yet complementary essences of nature; *wuxing*, which covers the five basic elements of the universe (wood, fire, earth, metal and water); *qi* or energy; and *xue*, the blood. This terminology purports to be concerned with disease and human health, but cannot be defined in terms of biochemical or biological facts — or indeed measured. And even the literal translation of these tenets into other languages is misleading.

The modern medical and scientific communities in China and elsewhere are highly critical of such mystical concepts, which are consequently becoming marginalized in China. "The medical practice of TCM is a process of trial and error, and concerns the understanding and control of herbs from the *Chinese Materia Medica*," says Daqing Zhang, director of the Center for History of Medicine at Peking University in Beijing. "The philosophical theories were created afterwards to provide the explanatory framework for the practices, and are used to win the patients' trust."

But thousands of years of history is a long time

GRACIA LAM

to cement a belief, and the theories of TCM still have their firm defenders, such as Boli Zhang, president of the China Academy of Chinese Medical Sciences (CACMS) in Beijing. “We believe in the jing luo [the meridian or energy pathways], but we have not found it yet,” he says. Nevertheless, even in the TCM community, there are fewer Chinese scholars who believe the TCM tenets literally. Indeed, there has been criticism from academics and the media in China, arguing that much of TCM and most of its theories are pseudoscience and that China should bid “farewell to traditional Chinese medicine”¹.

Despite these negative sentiments, the Chinese government has started to promote TCM by allocating large funds for research (see ‘One step at a time’, page S90). Influential figures in the TCM community, such as Boli Zhang and his deputy Baoyan Liu at CACMS, are pushing for the modernization of TCM while also emphasizing its unique qualities. These advocates insist that the traditional TCM practices should be preserved for as long as possible to better investigate their heritage. “The advantages and disadvantages of the two systems are the premise of the integration,” says Liu.

CLOSING THE GAPS

One practice in TCM that is adapting to science is diagnosis. Each patient receives a personal diagnosis based on a TCM syndrome or zheng, which is a characteristic phenotype of identifiable manifestations gleaned from general appearance, listening to and smelling the patient, feeling the pulse and asking questions. TCM diagnosis includes many symptoms considered less important in modern medicine, such as thirst, the tongue’s condition, whether the limbs feel cold, and mood. In contrast, modern disease diagnosis is based primarily on clinical signs such as temperature and blood pressure, pathological examination of individual organ functions and biochemical analysis of blood or urine.

“Chinese herbal medicine identifies and treats syndromes rather than diagnosed diseases,” explains Aiping Lu, director of the Institute of Basic Research in Clinical Medicine at CACMS. TCM syndromes are a finer level of classification than disease groupings in modern medicine. Lu’s group is collaborating with several Western research institutes on rheumatoid arthritis. In one study, his team used 18 different clinical TCM signs to classify rheumatoid arthritis patients into four subgroups, corresponding to presence of joint symptoms (tenderness, swelling and stiffness), cold pattern (joints and limbs feel cold to the touch, and patient is intolerant of cold), deficiency pattern (including weakness, dizziness, fatigue and nocturia) and hot pattern (including joints that feel hot, vexation, fever and thirst). Then all patients were randomized to receive either modern medicine or a TCM treatment. Each TCM-defined sub-group responded differently to treatment. “The results show TCM syndromes improved the efficacy of both TCM and modern medicine interventions,” says Lu.

Notably, patients with cold-pattern symptoms responded better to modern medicine, whereas those classified with a deficiency pattern benefited most from a TCM intervention². The next step is to analyse the patients’ metabolites to look for underlying biochemical differences between the different TCM syndromes, to see if there is a link between TCM diagnoses and modern medical definitions (see ‘All systems go’, page S87).

The composition of TCM medications is at odds with modern medicine too. TCM uses compound formulae (fufang) that contain several herbs thought to act in unison to restore what TCM practitioners call the patient’s ‘balance’. “It is an advantage of TCM that multiple

“Could Western treatments of chronic diseases be improved by insights from TCM?”

drug components can strengthen the therapeutic efficacy and attenuate the toxicity of major component(s) through drug–drug interaction *in vivo*,” says Wei Jia, co-director of the Center for Research Excellence

in Bioactive Food Components at the University of North Carolina at Greensboro.

Another advantage of a formula that acts on several fronts is that each ingredient does not have to be so potent. “If we use medium-active ingredients, which comprise the majority of compounds in nature, we have more potential drug candidates, achieve weaker side effects and have lower R&D costs,” says Zhimin Wang, a chemist at CACMS’ Institute of Chinese Materia Medica.

TCM practitioners formulate their herbal remedies according to a set of principles (peiwu), which organize ingredients in any fufang into four functional roles: sovereign, minister, assistant and envoy. The sovereign is the ingredient with major pharmacological activity. The ministers provide additive or synergistic activities. Assistants can either augment the pharmacological effect, detoxify or even counteract an excessively strong action. Envoys harmonize the whole recipe to ensure that all the substances in the fufang are compatible.

There is early evidence supporting the peiwu principles. A team led by molecular biologists Zhu Chen and Saijuan Chen at Jiao Tong University in Shanghai analysed the Realgar–Indigo *naturalis* formula — a TCM-based leukaemia treatment containing realgar, indigo minerals and the herb red sage root (danshen). They found that the arsenic in realgar could be described as the ‘sovereign’ as it attacked the main oncoprotein in leukaemia cells; indirubin, the active ingredient in indigo, worked as the ‘assistant’ to slow leukaemia cell growth; and tanshinone, from red sage root, served as ‘minister’ to help restore the damaged pathways that prevent leukaemia spreading. Indirubin and tanshinone also worked as ‘envoys’ to enhance cellular uptake of arsenic³.

“A TCM formula is structural, and using the ingredients should be like a military operation,” says Zhong Wang, a professor at CACMS’

Institute of Basic Research in Clinical Medicine, who has developed a concept that covers the science of TCM formulae — fangjiomics⁴. Fangjiomics separates biological networks into discrete, interactive modules that can be targeted separately but whose biological effect must be considered together. Zhong Wang argues that, compared with the Western model of single-target interventions, TCM combinations are systematic and well ordered.

THE LONG WAY AHEAD

The triumphs of modern medicine in the past 200 years have been most striking in preventing and curing infections and acute disease, and in pain relief. By contrast, progress in understanding and treatment of chronic and degenerative diseases has been slow. And it is these conditions, such as diabetes and Alzheimer’s disease, that are responsible for a large portion of the West’s soaring and unsustainable healthcare costs. Could Western treatments of these diseases, where there is no infectious agent but rather an internal imbalance in the immune system or other biological housekeeping systems, be improved by insights from TCM?

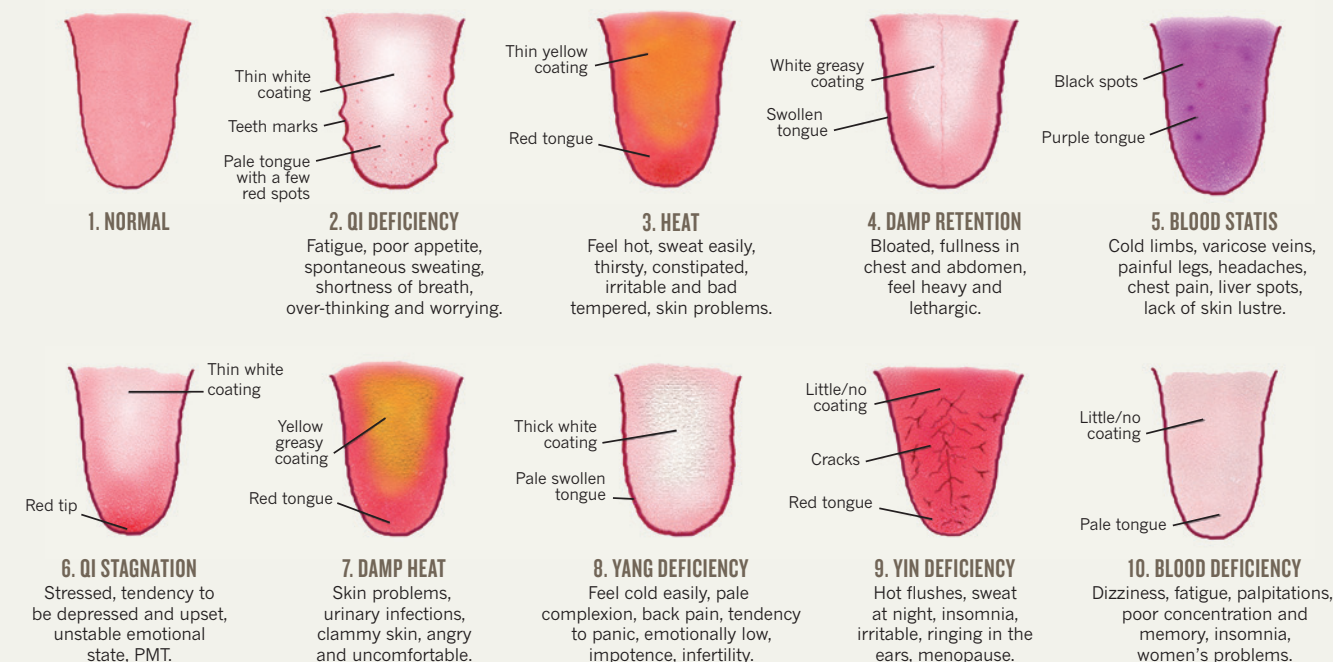
In modern medicine, chronic conditions are generally treated with prolonged administration of chemical drugs, which can give rise to long-term toxicity or even resistance. That’s what happens, for example, with many patients taking isosorbide dinitrate for chronic angina pectoris. The danshen dripping pill provides a better long-term curative effect, claims Jia. As it contains a mixture of ingredients, each individual component can be added at a lower, less-potent dose without compromising the effectiveness of the overall preparation. Moreover, TCM preparations tend to be cheaper than modern pharmaceuticals, he adds.

In addition to being a source of national and cultural pride, Chinese herbal medicines are the most promising source of new drugs for the Chinese pharmaceutical industry, which has always lived in the shadow of Western pharmaceutical giants. China is stepping up its efforts in both academia and the pharmaceutical industry to find safe and effective compound formulae that are acceptable to Western regulatory agencies. Most attempts have focused on isolating a single active ingredient from the TCM preparations. This reductionist, Western approach has led to only a few successes — including the antimalarial drug artemisinin and the leukaemia treatment arsenic trioxide.

So far, no drugs based on established TCM formulae have been approved in the United States or Europe. The nearest example is sinecatechins (marketed as Veregen by German biotech MediGene based in Martinsried), a cream made from a mixture of green tea extracts for the treatment of genital and perianal warts. Sinecatechins was approved in 2006 by the US Food and Drug Administration, and is the FDA’s first and only ‘botanical’ drug — approved on clinical results despite the fact that the active ingredients

TASTING NOTES

The tongue's condition is an important diagnostic tool in traditional Asian medicine.



and the mechanism of action are not known (see 'Herbal medicine rule book', page S98). However, the danshen dripping pill, manufactured by the Tainjin-based Tasly Group in China, has successfully completed phase II trials in the United States and could be the first botanical drug derived from the TCM repertoire.

"Pharmaceutical regulations in the West are developed according to the Western way of drug R&D, and they are not fit for evaluating TCM drugs," says Henry Sun, vice-president of Tasly Group. "Of course, TCM should be proved on the basis of scientific evidence." However, Sun argues that the science is skewed towards testing single agents targeted to single mechanisms (see 'The clinical trial barriers', page S93). "Besides the science, the other major challenge is convincing different groups in the West, from ordinary people to policymakers, to change their prejudices against TCM."

TOP DOWN OR BOTTOM UP?

The advent of -omics technologies that rapidly measure the entirety of the human complement of, for example, genes (genomics) or metabolites (metabonomics) — and to integrate these diverse data into a complete picture — has given rise to a new way of looking at medicine in the form of systems biology. For many TCM researchers, such as Aiping Lu and Zhong Wang, systems biology is potentially a way to understand TCM using Western scientific methodology.

"The major challenge of the integration of TCM and modern medicine is the translation from TCM experience and concepts into biochemical and biological meanings that Western

scientists can understand," says Jan van der Greef, a systems biologist at Leiden University in the Netherlands. "Systems biology is an ideal theory and analysis tool that can bridge the two systems."

Leroy Hood, president of the Institute for Systems Biology in Seattle, Washington, and regarded by many as the field's founding father, has introduced what he calls '4P healthcare': predictive, personalized, preventive and participatory. This concept, Hood contends, is the new paradigm of modern medicine in a systems biology era. 4P medicine focuses on the biochemical networks underlying health and disease, then aims to treat and prevent disease by identifying and countering perturbations in the biological networks — a concept highly reminiscent of the TCM philosophy. Thanks to systems biology, the gap between the two medical systems is starting to narrow. For example, recent advances in medical technology are allowing the application of new phenotyping technologies that can simultaneously characterize the multiple drug responses to dietary preparations, such as pu-erh tea. In a proof-of-concept study in humans, a team led by Jia quantitatively measured the absorption of pu-erh tea molecules, the output of gut bacteria metabolism, and the human metabolic response profile in the urine⁵. "The phenotyping strategy can further differentiate disease subtypes that are correlated to different TCM syndromes," says Jia. This, in turn, will "improve doctors' ability to personalize treatment and to predict an individual's response to a drug regimen".

But Jia and others are aware that the technology and methodology of systems biology are still immature. "We have to first develop network

models to simulate a pathogenesis: for example, the transformation of normal cells to cancer cells," says Jia. "The second step will be to model a multi-component agent targeting the multiple sites involved in the transformation process."

EAST MEETS WEST

The absence of scientific development processes and controlled clinical trials has held back the integration of traditional Asian medicine and modern medicine for centuries. Some of its concepts appear more magical than practical, and, without a physical basis, have resisted measurement and observation. But slowly these differences are resolving.

Much of the drive for integration will come from China and its neighbours. "We should build our own methodology to evaluate the unique features and efficacy of TCM," says Zhimin Wang. "There are so many possible ways of integration between TCM and modern medicine. We should keep our minds open."

The days of competition between these two systems could well be gone. "The two systems cannot replace each other," says Boli Zhang. "But instead they will fill each other out." ■

Peng Tian is a reporter for *The Economic Observer in Beijing*.

1. Zhang, G.-Y. *Med. Phil.* **27**, 14–17 (2006).
2. He, Y. et al. *Am. J. Chin. Med.* **36**, 675–83 (2008).
3. Wang, L. et al. *Proc. Natl Acad. Sci USA* **105**, 4826–4831 (2008).
4. Wang, Z. et al. *J. Clin. Pharmacol.* **51**, 1132–1151 (2011).
5. Xie, G. et al. *J. Agric. Food Chem.* **57**, 3046–3054 (2009).

PERSPECTIVE



All systems go

Systems science can provide guidance in capturing the complementary approaches to healthcare, says **Jan van der Greef**.

Various cultures have developed their own philosophies of science and, consequently, practices in medicine. The twentieth century has seemed to vindicate the Western approach, bringing huge advances in our understanding of physiology and biochemistry. This knowledge has fuelled the development of medicines and vaccines against countless diseases that had once wreaked havoc on humanity.

But in recent decades, the Western model has hit some turbulence. In particular, the concept of 'one disease — one target — one-size-fits-all' is shifting towards more personalized medicine tailored to individual patients, including the use of multiple therapeutic agents and the consideration of nutritional, psychological and lifestyle factors when deciding the best course of treatment. This shift in strategy has been most obvious in the prevention and management of chronic diseases such as diabetes and cardiovascular disease. The intellectual underpinnings for such a transition in medical practice are being laid in the discipline of systems science — and systems biology in the biomedical domain.

Systems science aims to understand both the connectivity and interdependency of individual components within a dynamic and non-linear system, as well as the properties that emerge at certain organizational levels. The relation to medicine is clear. Systems biology is particularly useful when it comes to describing homeostasis — the regulation of a system's internal environment to maintain a stable condition. In turn, the ability to cope with changing environments and stress is encompassed in the principle of allostasis — the physiological or behavioural changes required to stabilize the biological system.

The concepts and practices of systems biology align very closely with those of traditional Asian medicine. Consider the very idea of 'health'. The current World Health Organization definition of the term is based on a 1948 consensus: "A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." However, an emerging concept of health in the scientific literature describes an ability to adapt and self-manage in the face of social, physical and emotional challenges¹. This perspective has, of course, long been central to the concept of health in traditional Chinese medicine (TCM), which further includes spiritual fulfilment and a sense of individual well-being.

TCM is descriptive and phenomenological — it typically diagnoses patients using concepts based on the relationship between signs and symptoms, obtained through questioning, listening, palpation, visual inspection and smelling. In contrast, Western-style, modern medicine has mainly used single biomarkers to describe disease states, for example diagnosing type 2 diabetes by measuring glucose levels. But there is a growing realization in the West that single biomarkers are not enough. A better approach is to look at patterns of biomarker responses to a challenge. These data will provide insight into the resilience of allostatic mechanisms, and hence into a person's health, an approach not unlike the tenets of TCM.

In addition to giving Western medicine a basis for adopting some concepts of TCM, systems biology is also pushing the convergence from the other direction. Increasingly, TCM uses modern biochemical measurements and tools to refine or augment diagnostic descriptions. This is starting to facilitate the translation of TCM concepts into Western concepts based on biochemical, pathway or regulatory processes.

For instance, take diagnosis: the Sino-Dutch Centre for Preventive and Personalized Medicine in Zeist, the Netherlands, has conducted several studies that attempt to unify Eastern and Western diagnostic principles. In one such study of rheumatoid arthritis patients, selected according to Western (American College of Rheumatology) guidelines, TCM practitioners categorized patients into 'heat' or 'cold' pattern — based on the results of a questionnaire concerning joint issues, level and quality of pain, response to weather, and other symptoms such as fever

and thirst, as well as the results of a tongue and pulse inspection. A systems biology investigation of the two groups found statistically significant differences between them in the expression of genes related to apoptosis and metabolite profiles².

Systems biology can also provide insight into the multi-target pharmacology of herbal formulae. A metabolomic study investigated changes in lipid levels in transgenic mice with mild hypercholesterolaemia given either a herbal concoction or a known drug (for example, rimonabant, atorvastatin or niacin).

The study found that the herbal formula caused decreases in plasma cholesterol and triglycerides, and increases in high-density lipoprotein. How the herbal formula does this should help researchers pinpoint novel ways to treat metabolic disorders, especially those related to lifestyle³.

These early systems biology investigations suggest that the TCM method of qualitative

subtyping could be of use in deciding the course of treatment for patients in modern medicine⁴, and provide momentum for the move towards personalized medicine. Furthermore the concept of health promotion alongside disease management will help to improve the current system of healthcare. ■

Jan van der Greef is a principal scientist at the Netherlands Organization for Applied Scientific Research (TNO), professor of Analytical Biosciences at Leiden University and chairman of the Sino-Dutch Centre for Preventive and Personalized Medicine. e-mail: jan.vandergreef@tno.nl

1. Huber, M. *et al.* *Br. Med. J.* **343**, d4163 (2011).
2. van Wietmarschen, H. *et al.* *J. Clin. Rheumatol.* **15**, 330–337 (2009).
3. Wei, H. *et al.* *PLoS One* (in press).
4. Li, S. *et al.* *IET Syst. Biol.* **1**, 51–60 (2007).

The author declares competing financial interests: go.nature.com/tdy7sf



A flock of starlings relies on connectivity, dynamics and communication; these elements are embedded in TCM.



Traditional herbal remedies are typically prepared as tea.

MICROBIOME

That healthy gut feeling

Many ingredients in traditional herbal medicines cannot be absorbed by the human gut. Could our microbial inhabitants do for us what we can't do ourselves?

BY JAMES MITCHELL CROW

For an unassuming-looking little herb, *Panax ginseng* has quite a reputation. Pulling the plant from the floor of the forests where it grows wild reveals the origin of its fame — the hairy ginseng root, prized for centuries by Asia people for its medicinal properties. Dried and ground, ginseng is a key ingredient in the traditional herbal medicines of China, Japan and South Korea.

PETER FRANK/CORBIS

Investigation shows that ginseng is rich in a family of steroid glycosides, dubbed ginsenosides. Equally clear is the fact that these compounds have no direct effect on the human body because the gut cannot absorb them¹.

And yet, for many people, ginseng does have an effect. For although the human gut can't absorb the ginsenosides, some of the bacteria that live there can. After partly breaking down the compound for food, the bacteria excrete the remnants. It is these partly digested molecules that are used by the body and that have been proven to have a wide range of activity, from anti-inflammatory to anticancer effects.

The human gut is home to trillions of individual microbes representing thousands of species of bacteria and non-bacterial organisms called archaea². The exact membership of this highly complex ecosystem, known as the microbiome, varies from person to person. Indeed, according to microbiologist Liping Zhao at Jiao Tong University in Shanghai, China, for around one-fifth of the population, ginseng will have no health benefit because the person doesn't have the right gut microbes to break it down.

Our microscopic passengers have long been suspected of being active participants in a mutually beneficial partnership. Integral to this picture is the interplay between gut bacteria and health. The global rise of chronic health conditions, ranging from obesity and diabetes to bowel disease and cancer, is increasingly being linked with perturbations in gut flora. And while modern medicine is struggling to tackle such multi-component diseases, the ancient medical philosophies and practices of Asia — particularly those of traditional Chinese medicine (TCM) — could offer an alternative approach. TCM's reliance on complex mixtures of compounds, and its philosophy of treating the human body as a whole, complete system that needs to be balanced, matches up well with the synergistic properties of the gut microbiome.

DUAL BENEFIT

Herbal medicines can affect health via the gut microbiota in two ways. Like ginseng, a host of herbal medicines are known to take effect only after being processed by bacteria in the gut. The list includes the dried fruits of *Gardenia jasminoides*, containing the compound geniposide, which is converted by gut microbes into its active form, genipin, another anti-inflammatory and anticancer compound. Similarly, the root of the liquorice

plant, *Glycyrrhiza glabra*, contains glycyrrhizin, which can be processed by microbes into 18 β -glycyrrhetic acid — effective in the treatment of peptic ulcers, as well as having antiviral and antifungal activities³.

In the other type of interaction, certain ingredients in herbal medicines influence the balance of bacterial species living in the gut. For example, extracts from the *Ginkgo* leaf have been shown to increase the abundance of beneficial bacteria such as *Lactobacillus* and bifidobacteria in the gut. These bugs have been linked with a number of health benefits in the human host; in particular, they can modulate the immune system in ways known to reduce the risk of autoimmune diseases such as diabetes mellitus type 1 (ref. 1).

That at least some of the ingredients in herbal medicine exert their biological effects through interactions with our gut microbiota is not a new revelation. As long ago as the 1950s, Wei Xi, a microbiologist in the Dalian Medical University, China, proposed that the key to understanding the action of herbal medicines in the body was to consider their interaction with the microbes in our gut.

“People have appreciated for a long time that we carry this enormous community of microbes around with us,” says George Weinstock, a geneticist at Washington University in St Louis, Missouri. What has long limited our understanding of the role that the gut microbiota play in health was the overwhelming number of organisms present. The only realistic approach is to sample the whole bacterial community at once, says Weinstock. “It has only been in the last five years or so that the power of DNA sequencing instruments has got to the point where you can tackle that kind of project,” he says. Several large projects have already begun to chip away at the problem, including the US National Institutes of Health’s Human Microbiome Project, co-directed by Weinstock, and the European Commission-funded Meta-HIT project.

The next big step will be to link big-picture changes in microbiome makeup with particular patterns of disease, says Weinstock. Researchers are already starting to spot links between certain chronic diseases and the absence or presence of specific bacteria. For example, people with Crohn’s disease, a painful autoimmune disease affecting the bowel, tend to have low levels of a bacterium called *Faecalibacterium prausnitzii*, which is, therefore, suspected to play a protective role⁴. Another bacterium, a particular form of *Escherichia coli* called adherent invasive *E. coli*, is more prevalent in people with Crohn’s disease, raising the prospect it might be one of the bad guys⁵.

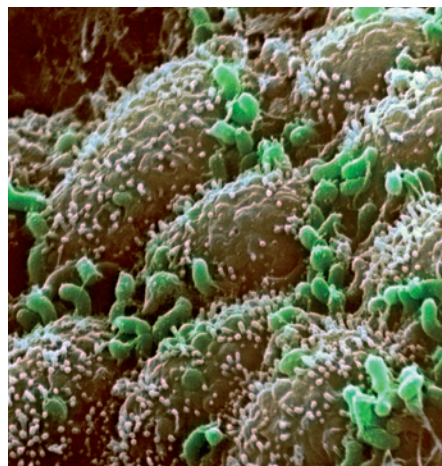
However, there is more to disease development than a single causative bacterium, says Weinstock. “It’s about the conversation between you and your microbes, and when that

conversation gets out of whack (for reasons that we don’t yet know) and if you have the right susceptibility genes, then disease can occur.”

It is this multifaceted nature of chronic diseases that makes them so hard to treat. “Western medicine is extremely good for treating acute diseases,” says Jan van der Greef, an analytical scientist from the Netherlands Organization for Applied Scientific Research (TNO) based in Delft, who helped establish the Sino-Dutch Centre for Preventive and Personalized Medicine in Zeist, the Netherlands. But the cell-based approach to medicine in the West is less effective for other kinds of ill-health, he says. “If you move towards chronic illnesses — or even to disease prevention and health promotion — our current way of thinking is really limiting what we can achieve.”

TROUBLE ON THE WESTERN FRONT

Could it be that the traditional Eastern approach to health is better equipped to tackle chronic disease? “Because these complex diseases have multifocal problems, no single drug can treat them,” says Jeremy Nicholson, a biochemist at Imperial College London. “Chinese medicine is a polypharmacy, with multiple synergistically active compounds in the mixtures; the reason some of the medicines probably work is that they drug multiple targets at the



Bacteria (green) settled on the small intestine.

same time.” And in the gut, there are thousands of potential targets.

In fact, as far as chronic diseases go, modern medicines could be part of the problem, Nicholson adds. “Antibiotics don’t just kill bad bugs, they kill good bugs as well.” This disruption can still be detected in the gut microbiome at least two years after a patient has completed a course of antibiotics⁶.

Nicholson studies the interactions between gut microbiota and health using a whole-body systems biology approach that he invented called metabonomics — a systemic-wide version of metabolomics. “What we try to sample in metabonomics is the systemic response to some sort of intervention,” he says. The team uses analytical chemistry techniques such as

nuclear magnetic resonance (NMR) to map all of the metabolites — the chemical by-products of metabolism — in an individual’s blood, urine or stool sample, thereby capturing the cell’s metabolic output. By comparing the results before and after a treatment, metabonomics provides a read-out of the whole body’s response to a drug — whether that drug has worked directly on a cellular target or indirectly by interacting with the gut microbiome. This systemic view is the key difference between metabonomics and the standard scientific approach to examining metabolites using cell cultures.

For the past few years, Nicholson has been collaborating with Zhao, who is a leader in researching the interaction between herbal medicines and the gut. Zhao believes that metabonomics is uniquely placed as a tool for understanding how herbal medicines work, because it captures the whole-body response to what is, by intent, a whole-body treatment.

“We’ve done some work showing that herbal medicines produce notable shifts in gut microbial metabolism, and those shifts can be quite stable over quite long periods of time,” Nicholson says. Giving people chamomile, for example, changes the metabolite make-up of their urine. These effects remain even when the herb is no longer taken, implying that the change is caused by a lasting shift in gut microbe metabolism. “There is almost certainly a strong connection between microbiome activity and the activity of traditional Chinese medicines.” Not that herbal medicines were deliberately designed to work this way, he adds. “For TCM practitioners, it’s complete news to them.”

Zhao says that interactions with the gut microbiota could ultimately prove to be one of the main ways in which herbal medicines act on human health. “In Chinese medicine, many ingredients just pass through the gut, they don’t get into the bloodstream.” And yet some of these ingredients are known to have an effect, he says. So, much like ginseng, “it is most likely they work by changing gut microbiota.”

Given the emerging links between gut microbiome and human health, it is these types of glimpses that are inspiring scientists such as Nicholson to investigate traditional herbal medicines as a new way to treat chronic diseases. “If we can start to unravel how TCM works,” he says, “it might offer a completely new horizon on how you drug the human body.” ■

James Mitchell Crow is a freelance science writer based in Melbourne, Australia.

1. Zhang, X. *et al.* *Mod. Tradit. Chin. Med. Mater. Med.* **13**, 202–212 (2011).
2. Jia, W. *et al.* *Nature Rev. Drug Disc.* **7**, 123–129 (2008).
3. Kim, Y. S. *et al.* *J. Microbiol. Biotech.* **18**, 1109–1114 (2008).
4. Sokol, H. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).
5. Darfeuille-Michaud, A. *et al.* *Gastroenterology* **115**, 1405–1413 (1998).
6. Jernberg, C. *et al.* *ISME J.* **1**, 56–66 (2007).

➔ NATURE.COM

For some of the latest research on metabonomics: go.nature.com/ix34rg



In the process of purifying and testing the antimalarial agent artemisinin, phytochemist Youyou Tu realized that the age-old Chinese technique of boiling and high-temperature extraction destroys the artemisinin contained in *Artemisia annua* (*qinghao* in Chinese), a herb used in traditional Chinese medicine (TCM) for hundreds of years. So she redesigned the process for isolating artemisinin, performing it at lower temperatures (60 °C) and using ether as a solvent. Tu also removed a harmful component of the herb that did not contribute to antimalarial activity, discovered that the active ingredient is present only in the leaves, and worked out the best time to harvest it. Since its initial isolation in 1972, artemisinin and its derivatives have become the leading treatments for one of the world's most deadly diseases and saved countless lives. On 12 September 2011, Tu, who has since retired as a research fellow from the China Academy of Chinese Medical Sciences (CACMS) in Beijing, won the Lasker–DeBakey Clinical Medical Research Award for her pioneering work on this agent.

GRACIA LAM

The other classic example of a modern medicine that originates in TCM is arsenic trioxide, which was approved by the US Food and Drug Administration for use in leukaemia in 2000. Tingdong Zhang, a research fellow of Harbin Medical University in China, made a unique contribution to the development of this drug by separating a toxic component from the therapeutic compound. Tingdong Zhang used animal toxicological experiments to define the maximum safe dose, and changed the route of administration to intravenous delivery to maximize the effect.

“The discoveries by Tu and Tingdong Zhang are great examples of how to develop drugs from TCM,” says Bai Lu, vice-president of biology at GlaxoSmithKline (GSK) China in Shanghai. “Consciously or not, they were actually using modern scientific approaches for drug discoveries.”

Beyond the utility of these drugs as lifesavers, their development raised hopes in China that, with the help of modern science, TCM could become an equivalent — or perhaps even a more powerful counterpart — to modern medicines. But, despite decades of research and billions of yuan, TCM is still considered to be part of complementary or alternative remedies in most Western countries, its ingredients hard to define, and its efficacies difficult to evaluate.

ACTIVE INGREDIENTS

For decades, European and US regulatory agencies held the view that a drug must be either a highly purified or synthetic agent. Traditional medicines could not qualify under this definition because a TCM preparation is a concoction rather than a single compound, and the chemical structures of the active ingredients are undefined. Consequently, when it came to assessing a traditional medicine for efficacy, most effort went into identifying the principal agent within a herb or mixture of herbs and purifying it as a

MODERNIZATION

One step at a time

The repertoire of traditional Chinese medicine could offer rich pickings for modern drug developers, but researchers must first define and test herbal concoctions.

BY ZHIGUO XU

single chemical compound. However, “academic scientists and the pharmaceutical industry have not been very successful at isolating the active substances in TCM preparations”, says Bai Lu.

Others share Lu’s dour outlook. “Few active ingredients are extracted from TCM herbs by modern scientific approaches,” says Yi Rao, dean of life sciences at Peking University in Beijing. “Those extracted effectively have not been demonstrated to be clinically successful.” Indeed, he says, “artemisinin and arsenic trioxide are perhaps the only two examples”.

This process is not helped by the insistence of many TCM researchers on using traditional methodologies — such as water decoction, immersion, reflux extraction and distillation — to process herbs. “It is totally ridiculous,” says Rao, who is also an evaluator for the big-budget Innovative New Drug Development and Manufacturing project, which was launched in 2008 with more than 10 billion yuan (US\$1.6 billion) of public funding. The aim is to develop new TCM drugs by applying modern research and development techniques to TCM. Rao criticizes TCM researchers who adhere to traditional ways. “They think it’s too slow and troublesome to extract the active ingredients by modern scientific approaches, but in fact this was how Tu and Tingdong Zhang did it”.

A major drive of the central Chinese government is to sponsor projects that use modern drug discovery technologies to characterize the active parts of TCM formulae. For example, the 30 million yuan (\$4.7 million) Herbalome project, initiated in 2008 by scientists at Dalian Institute of Chemical Physics (DICP), part of the Chinese Academy of Sciences, is compiling a database of active agents in herbs from popular TCM recipes¹. Researchers use liquid chromatography and other modern methods of separation to identify the components of a herb, then evaluate each one using high-throughput screening, toxicity testing and randomized clinical trials to determine both its active ingredients and toxic components. More than 10 scientists are involved in the project, which is set to last for 15 to 20 years.

Xinmiao Liang, a chemist at DICP and head of the Herbalome, is confident about its prospects.

“We can analyse 500 samples every day, and initial results are very good,” he says. “We have already extracted the active ingredient from one herb and identified its chemical structure.” Liang is applying for international patents, conducting animal experiments and writing papers about this newly identified analgesic compound, although he is yet to publish his findings. Liang hopes the Herbalome will uncover the next artemisinin or arsenic trioxide.

COMPLEXITY

But within China, the Herbalome project has its sceptics. “Japanese scientists have tried to extract active ingredients from TCM herbs in past decades, but failed,” says Bengang Zhang,



Youyou Tu is bringing TCM into the modern clinic.

a researcher at the Institute of Medicinal Plant Development of the Chinese Academy of Medical Sciences (CAMS) in Beijing. Part of the problem is that standard laboratory technologies do not always match up well with TCM analysis. For example, Liang acknowledges that one of the major tools used by his team as part of the Herbalome — high-throughput screening — is not ideally suited to the task. That’s because TCM concoctions are mixtures of multiple active compounds, and a typical Chinese medicine is intended to hit multiple biological targets. Hence, even if a high-throughput screen shows a strong effect, it cannot identify the active ingredient — especially if it is a trace component. “In most cases, you don’t know what the targets and mechanism are,” says Bai Lu. “Without that, you

can only identify and isolate compounds based on phenotypes, which is extremely difficult”.

According to Bengang Zhang, the failure of the Japanese researchers was not because they were using poor methods or technology. “The instruments in Japan are sophisticated enough,” he says. The issue, in his view, is that both the human body and the formulae in the TCM repertoire are complex systems. Trying to simplify both, by extracting one active ingredient to hit one biological target, is a recipe for futility.

The difficulties have not deterred scientists keen to explore traditional medicines and untangle their many components. “This is actually an advantage of TCM,” says Bai Lu. Many common diseases, including diabetes and hypertension, are polygenic and their pathology follows multiple mechanisms. Alzheimer’s disease, for example, involves accumulation of toxic proteins (amyloid- β and tau), neuronal death, loss of function in the synapses, and inflammation. A drug that hits a single mechanism, no matter how hard, might not work. Bai Lu adds: “The multiple components in TCM formulae offer a unique opportunity to attack multiple disease-causing mechanisms simultaneously — provided we know what they are and how to test them clinically”.

Taking advantage of complexity is one of the goals of Zalicus, a drug developer in Cambridge, Massachusetts. “Biological systems have evolved over millions of years to be able to resist a single insult and they have become very robust,” explains Margaret Lee, vice-president of research at Zalicus. “There is a redundancy and complexity in biology”.

Zalicus’s approach is to take existing molecules with known mechanisms of action, and combine them to look for synergistic activity. For example, Zalicus, in partnership with Sanofi Aventis of Paris, France, is developing Prednisporin — a combination of the glucocorticoid prednisolone acetate and the immunosuppressant cyclosporine A — for the treatment of inflammatory eye diseases.

However, working the other way around — starting with a complex herbal mixture and trying to “deconvolute the synergy” — is tough, says Lee. “With a herbal medicine, there might

FUTURE CHINESE TREASURES?

Several drugs are aiming to become the first authentic TCM-derived formula to get approval in the West.

Drug name	Manufacturer	Active ingredients (if known)	TCM formula	Diseases to treat	Status
Danshen dripping pill	Tasly Pharmaceutical	Danshensu, protocatechuic aldehyde, salvianolic acid B	<i>Salvia miltiorrhiza</i> , <i>Radix notoginseng</i> , Borneol	Cardiovascular disease (especially chronic angina)	FDA phase II clinical trials
Kanglaite injection (KLT)	Zhejiang Kanglaite Group	Unknown	Coix seed oil (with soybean phospholipids and glycerol as excipients)	Cancer	FDA phase II/III clinical trials
Xuezhikang capsule	Beijing Peking University WBL Biotech Co.	Various natural statins, unsaturated fatty acid, ergosterol, alkaloids	100% Red yeast rice (Hong Qu), obtained through fermentation of with <i>Monascus</i> strains.	Coronary heart disease caused by hyperlipidaemia and atherosclerosis.	FDA phase II clinical trials
Fuzheng Huayu capsule	Shanghai University of TCM	Salvianolic acid B, Adenosin	<i>Salvia miltiorrhiza</i> , peach seed, pine pollen, <i>Gynostemma pentaphylla</i> , <i>Schisandra chinensis</i>	Liver fibrosis	FDA Phase II clinical trials
Nongsuo Danggui Pill	Lanzhou Foci Pharmaceutical		Angelica root	Thrombosis, blood lipid levels and pain	Applying for registration in EU



CORBIS / HUGH LAU / MAYWAY CORP

From rural farming to the lab: herb farming standards were introduced in China in 2002 to help standardize the raw ingredients of herbal remedies.

be 10–15 different components in there.” A researcher would need to isolate the components, purify them, then add them back in various combinations until the entire effect of the mixture is recapitulated. And that is supposing it is possible to identify the components in the first place. “With a herb, where it’s grown and the environmental conditions can have an impact on the level of components,” she says. It would take repeated attempts to ensure all active ingredients have been isolated. “In theory, it should be possible. But in practice it can become very difficult.”

QUALITY ASSURANCE

Bai Lu says that consistency is an important challenge. For each recipe, “some sort of standardization criteria must be established to control batch-to-batch variation”, he says. Bengang Zhang agrees: he helped draft the TCM Good Agricultural Practice (GAP) guidelines for the State Food and Drug Administration (SFDA), the Chinese equivalent of the US Food and Drug Administration (FDA). He contends that poor agricultural practices lead to inconsistent quality of TCM products, severely affecting stability and therapeutic effects². Bengang Zhang proposed that these problems can be addressed in part through a GAP approach, which standardizes various basic drug production processes, including the cultivation, collection and processing of TCM herbs. These techniques should improve the quality of the herbs and bring TCM in line with international practices.

GAP for herbal products was introduced in China in 2002. Although it is not mandatory for herbs, being GAP certified can promote the visibility of a manufacturer. In the eight years to 2010, the SFDA certified 99 production bases (about half of the total number), which ranged in size from around 13 hectares to hundreds of hectares.

There are still problems that GAP has not solved. The rural areas where most herbs are grown remain underdeveloped, and many young people leave the countryside to work in the cities.

The agricultural land is typically owned by the drug manufacturer, and even though the companies provide training, the GAP guidelines will not be properly implemented without an adequate labour force. Basic research into important herbal production technology is weak, Bengang Zhang says. Moreover, he points out, “the government has not yet enacted any provisions to enforce implementation of GAP for TCM herbs”.

NON-TRADITIONAL TRIALS

But even with standardized raw ingredients and consistent products, measuring the clinical efficacy of TCM formulae is a challenge. Multiple active compounds hitting multiple targets could also mean multiple criteria for judging whether the medication works. While scientists worldwide are becoming increasingly interested in the evaluation of TCM products based on clinical trials, few researchers know how to do it right. It is clear that double-blind, randomized clinical trials (RCTs) are needed.

“Some sort of standardization must be established to control batch-to-batch variation.”

to measure multiple clinical endpoints based on multiple disease mechanisms, especially when you cannot measure them separately.”

Aiping Lu, researcher into integrative medicine in the Department of Disease and TCM Pattern Correlation Research at CACMS, says that more than 100 clinical trials on TCM formulae were conducted in China from 2006 to 2010 — most of them RCTs — and that more are coming soon. “A large amount of TCM clinical research papers will be published in the next 2–3 years,” he predicts. Indeed, several drugs developed from TCM recipes are currently in clinical trials, registered with agencies in Western

countries (see ‘Future Chinese treasures’).

But the widespread adoption of such robust, modern testing techniques requires a significant change from today’s situation, says Bengang Zhang. “Only a few dozen new TCM drugs are approved by the SFDA each year. Some approved TCM formulae have to be reviewed again, because the registration and approval processes have become stricter.” He adds that many products “won’t be able to enter into clinical trials even if they are successful in animal experiments”.

Indeed, despite the popularity of TCM, there is clearly a long way to go before its drugs pass the rigorous scientific standards of Western pharmaceuticals. According to the SFDA’s Annual Report on Drug Registration Approval, only 81 TCM products were approved in 2010 — 8% of all drugs, and slightly lower than the figure for 2009. The overwhelming majority (89%) of drugs that were approved in China in the same year were synthetic or purified chemicals.

Modernizing Asian medicines will require more than just applying modern tools and scientific techniques to ancient practices. Simply extracting one active ingredient from a herb or herbal concoction and then trying to find its biological activity has produced few positive results in the past and is unlikely to work effectively in the future. The process will require a better understanding of how multiple ingredients act in synergy, and what effect they can have on multiple targets. If TCM is to take its place in the modern medicine cabinet, then it must develop modern ways to prove itself. “TCM is a great treasure,” says Bai Lu. “But the only way to bring its full values to society is using modern science and drug discovery technologies.” ■

Zhiguo Xu is a freelance journalist based in Beijing. Additional reporting by **Michelle Grayson**.

1. Zhang, X. et al. *Anal. Bioanal. Chem.* [Epub ahead of print] (2011).
2. Zhang, B. et al. *Planta Med.* **76**, 1948–1955 (2010).

Protecting China's national treasure

Applying modern intellectual property standards to ancient medicines.

BY JESSIE JIANG

The very nature of traditional Chinese medicine (TCM) poses serious challenges in patenting intellectual property. Unlike modern medicines, which are based on newly discovered, developed or isolated chemical compounds, TCM prescriptions typically consist of mixtures of well-known plant or animal extracts. Moreover, these recipes are public knowledge, having been recorded in ancient books, and are therefore ineligible for patents in areas where a compound's novelty is a prerequisite.

Even if a TCM prescription secures a patent — possibly after being safeguarded as a family secret, or as a modification to an established recipe — there are loopholes that allow the patent to be bypassed. According to TCM theories, each prescription has a *fangyi*, which is a description of the symptoms and an explanation of how they should be treated. A TCM prescription can maintain its *fangyi* even when one or several of its ingredients are replaced by similar items. For example, different species of ginseng can be used interchangeably in TCM prescriptions. This gives TCM companies leeway: under current Chinese patent law, even slight variations of a patented prescription can qualify for a new patent.

"Adding or removing an ingredient sometimes doesn't much change the efficacy of a compound TCM prescription," says Kaixian Chen, president of the Shanghai University of Traditional Chinese Medicine. This imprecision, he says, "not only leaves a lazy shortcut for patent applicants, but also undermines existing TCM patents."

Indeed, China's patent law for TCM is such that it even allows for a change in packaging design to qualify for a patent, says Xiaoting Song, a law professor at Tongji University in Shanghai who specializes in TCM intellectual property rights. "The intellectual property rights protection offered by the current patent system is weak at best," he adds.

And while TCM continues to be a source of pride for the Chinese, lax intellectual property rights create an opportunity for foreign-made medicines to enter China. According to reports

from Xinhua, China's state news agency, TCM imported from Japan, South Korea and Germany accounts for approximately 30% of the Chinese market.

China only started awarding patents for TCM products in the 1990s, and there are still no specific standards that govern whether a particular TCM formula is eligible for a patent — the same rules also govern food recipes, for example. "There is obviously an essential difference between a TCM formula and other compound products," says Song. "The concept of *fangyi* is unique to TCM prescriptions and should be protected by the patent law." Although China's State Intellectual Property and Trademark Office is considering new standards specifically for TCM, any such policy would take years to

"Protection offered by the current system is weak at best."

be approved by the central government. Meanwhile, TCM patents for new products with little added value continue to be awarded. "My sense is that at least half of all

TCM patents in China are for packaging design rather than for the medicine itself," says Song (there are no official figures).

Despite loopholes in the patenting of TCM formulae, some global pharmaceutical companies are keen to develop new medicinal products and are expanding their R&D efforts into the realm of traditional medicine. "Collaborations are becoming increasingly necessary when it comes to pharmaceuticals," says Song. "Even back in the 1960s, the discovery of the antimalarial artemisinin was a joint effort that involved more than 500 researchers from 60 research institutes in China."

International pharmaceutical companies submit a small fraction of TCM patent applications in China — about 3% according to the Intellectual Property and Trademark Office. But this is set to grow. "As the pharmaceutical industry gets more competitive, TCM represents a very promising approach," says Bai Lu, vice-president of biology at the Shanghai R&D centre of GlaxoSmithKline (GSK), the London-based global

pharmaceutical company. "We are very optimistic about the TCM market." Indeed, GSK is rapidly expanding its TCM team in Shanghai, and has established collaborations with several local institutions. When it comes to new pharmaceutical agents, Lu considers the current patenting system in China to be industry friendly — companies can essentially apply for two patents for each new compound derived from a TCM formula: one for the extraction and isolation method, and a second for the end product.

Securing a patent for a new TCM product can bring great financial reward. For example, the danshen dripping pill, manufactured by Tianjin-based Tasly Pharmaceutical, is a patented cardiovascular drug that is a more concentrated and faster-releasing version of the generic danshen tablet, made from the dried root of *Salvia miltiorrhiza*. Since its launch in 1994, revenue from the danshen dripping pill has grown by more than 10% each year on average, reaching US\$203 million in 2010 — accounting for 28% of Tasly's total revenue.

At the same time as encouraging modernization of TCM, the Chinese government is also taking steps to preserve the traditional practice. The Chinese Ministry of Health recently approved a draft of new regulations that, if approved by the central government, will officially categorize TCM as 'traditional knowledge', a status defined by the World Intellectual Property Organization, a United Nations agency headquartered in Geneva, Switzerland. The aim of the new ministry of health regulation is to protect historic TCM prescriptions. For example, the Sijunzi decoction, a 900-year-old Chinese recipe that mixes four plants, would be recognized as traditional knowledge under the new regulation, and its derivative recipes would not be considered patentable. "TCM is our national treasure," says Song, who was a member of the drafting committee for the regulations. "It will give us the edge when competing on the world stage, and we can't afford to lose that." ■

Jessie Jiang is a journalist based in Beijing.



GRACIA LAW

Enter the clinic

The editor of Nature China reports on his first visit to a traditional Chinese medicine practitioner to find out how this ancient practice is dispensed in the twenty-first century — and to see if anything can be done to relieve his back pain.

BY FELIX CHEUNG

The taxi rounded a corner and I wound down the window to let in some fresh air. Outside I could see vibrant-coloured spittoons, hear the sharp clang of metal being worked and smell the scent of washed clothes hanging from ancient balconies. I had arrived in the old Yuexiu district of Guangzhou in south China. Although much of the district has taken on the modern sheen of corporate business, Yuexiu still contains enclaves of bygone times.

I was here to see if my persistent back problem could be helped. The past five years of office work and sitting at my desk had taken their toll. I had days when I woke up but could not get out

of bed, and times when I tried to put on my socks but could not reach my feet.

Blood tests, X-rays and magnetic resonance imaging — everything at my Western-trained doctor's disposal had failed to identify the root of my pain. He assured me that I did not have arthritis, cancer or any inflammatory diseases, and that the problem was likely to be mechanical. All he could give me were painkillers — and antacids to quell any upset stomach that the painkillers caused.

Over the past few months my back pain had worsened. Walking or straightening my left knee would induce a tingling sensation at the back of my leg that radiated up to my hip and down to my toes. It got to the point where I started looking for alternative treatments to help me return to normal life. I researched medicine establishments online and asked around for recommendations, and found one

place that sounded suitable, even though it was 170 km northwest of my home in Hong Kong.

THE OLD AND THE NEW

The taxi stopped in front of Guangdong Provincial Hospital of Traditional Chinese Medicine. This was my first visit to a traditional Chinese medicine establishment. Even from the street, this hospital looked unlike other modern hospitals I been to before. Hanging on the wall near the entrance was a large LED display. Local people were standing below it, performing a series of slow movements and stretches in time with instructors on screen. They were doing *baduanjin*, a form of medical qigong, to strengthen their *qi*.

Guangdong Provincial Hospital actually provides integrative treatment, combining TCM and modern medicine. The hospital also claims to be the largest and busiest integrative



I had never tried acupuncture before, so my TCM doctor, Xiuhua Chen, gave me a trial run by inserting a needle behind my ear.



Above: The pharmacy at the Guangdong Provincial Hospital stores its Chinese herbs in traditional wooden cabinets. Left: Women practising *baduanjin* in front of the Guangdong Provincial Hospital.

hospital in southern China: more than 5.6 million patients visited in 2010 — around 15,000 patients each day.

TRADITIONAL APPROACH

Once inside, my visit began on a familiar course to any visit to a medical facility. I filled in a registration form and received a smartcard to store my medical records. Then a nurse led me to my TCM doctor — Xiuhua Chen, who is also director of the Traditional Therapy Center. I told Chen of my medical history and described my symptoms. She asked me about my back problems as well as other questions related to how well I was sleeping and about my bowel movements. Then Chen took a good look at my tongue and felt my pulse.

But then the Asian approach to medicine became apparent. Chen concluded that the tingling sensation in my leg was caused by my back problem, and that I had *xinhuo* or 'heart fire' — a condition that causes restlessness, insomnia and oral ulcers (which I also suffered from). Chen offered me acupuncture to alleviate my pain, moxibustion and cupping to increase blood flow to my back, and bloodletting to clear the heart fire. She also reiterated what my other doctor had told me — that I might need to resort to surgery if my symptoms persisted.

Because I had never had acupuncture, I was unsure of what to expect — so Chen gave me a trial run. I sat down in a chair and she inserted

a very thin, solid, stainless steel needle, about 3 cm long, into my head — just behind my right ear. When I stood up, I could feel that the tingling sensation had disappeared.

I was astonished: the needle insertion was fast and painless, and its effect instantaneous.

The trial over, I followed Chen to a private room where I removed my clothes and lay down on my side on a treatment table. Chen inserted 2 needles into my head, 16 more into my back and 3 into my right ankle. This time, however, I could actually feel the needles enter my body. I told Chen where I felt the most pain and she

“Finally it was time for the bloodletting and the nurse inserted more needles into the backs of my thighs.”

adjusted the depth and location of the needles. Chen then left the room and a nurse entered, pushing a trolley loaded with more needles, mugwort sticks, glass cups, and an alcohol burner. The nurse lit a mugwort stick and started circling it behind my back (which was still bristling with needles). Because I was facing away from the nurse, I could not see what she was doing, although I could smell the woody, spicy aroma of the burning mugwort, sense smoke against my back and feel heat through the needles. The moxibustion took around 15 minutes in total.

The nurse then pulled out all the needles,

wiped my back with an alcohol solution, and started the cupping. By heating the air in the cup, each about 10 cm tall, and pressing it against my back, she created a vacuum that held it in place. I found the heat and the suction of the cups quite comforting, and a lot more pleasurable than acupuncture or moxibustion.

HERBS FOR HOME

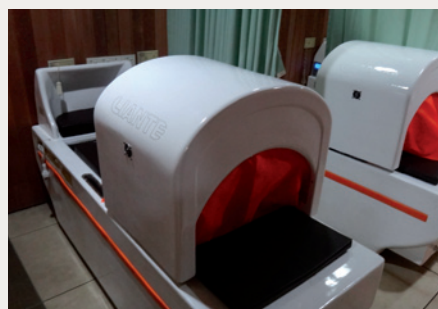
Finally, it was time for the bloodletting. The nurse inserted more acupuncture needles into the backs of my thighs and covered them with heated cups. Within 5 minutes, blood began to trickle from the base of each needle.

My treatment complete, the nurse led me to the TCM pharmacy to receive my medication. From the waiting area I could see the walls lined with traditional wooden cabinets where the herbs are stored. The herbs handed to me (to prepare at home — although they can do it for you on-site) were vacuum-packed, with the name, origin and manufacturing date clearly printed on the label. All in all, my bill came to 300 yuan (about US\$50).

I walked happily away with my leg feeling much better. As I left the hospital, I saw the women still doing their *baduanjin*. I began to wonder if maybe I should start practicing *baduanjin* too. ■

Felix Cheung is editor of *Nature China* in Hong Kong.

Below: The Traditional Therapy Centre at the Guangdong Provincial Hospital has a modern reception. As this was my first visit, I had to fill in a registration form. I was also given a smartcard to store my medical records. Right: I was prescribed traditional herbs to treat my *xinhuo* (heart fire), which I was to take home and prepare into a type of tea. The herbs were vacuum packed and clearly labelled with information on the origin, net weight, and production and expiry dates.



Left: Treatment beds in the Traditional Therapy Centre set up for fumigation, which is a bit like aromatherapy. The machines blow herbal steam onto the body.

JAPAN

Will the sun set on Kampo?

The practice of traditional medicine in Japan includes many modern techniques but faces numerous challenges — including political pressure from China.

BY ICHIKO FUYUNO

Japan's traditional herbal medicine, called Kampo, has become thoroughly integrated into the country's healthcare system. Of the thousands of Kampo formulae available, 236 are officially registered by the Japanese ministry of health and produced in pill or powder form. More than 80% of medical doctors in Japan currently prescribe Kampo, and two-thirds of the officially registered formulations are covered by national health insurance.

But Kampo — imported from China 1,500 years ago — is under pressure. Despite growing support among the Japanese public, Kampo is eclipsed by Western medicine, which accounts for 98% of total pharmaceutical production in Japan. Kampo also faces fierce competition from the growing popularity of traditional Chinese medicine (TCM).

Part of Kampo's problem is institutional: unlike China and South Korea, where governments promote traditional medicine, Japan has neither government departments nor public institutes dedicated to Kampo. The Japanese government spends about one-tenth as much on traditional medicine research as South Korea, which is only a fraction of the size of China's investment. And the difficulty of accumulating scientific evidence in support of traditional medicines makes the government wary of strengthening its support.

HISTORY OF KAMPO

In the sixth century, TCM spread to Japan and started to develop independently. Kampo medicine bloomed in the eighteenth century when Japan closed itself off from contact with most foreign countries.

After Japan reopened its borders in the late nineteenth century, the government embarked on a strategy of Westernization, and Kampo was all but replaced by modern medicine. Since then, Kampo has survived on the sidelines of society. In 1976, facing strong political pressure from Taro Takemi — then president of the Japanese Medical Association and a Kampo advocate — the health ministry decided to include Kampo in the national health insurance. But this policy

did not require clinical trials to prove the efficacy of Kampo treatments. Indeed, until recently, little importance was placed on the development of Kampo; only a handful of places, including Toyama Medical and Pharmaceutical University (now known as the University of Toyama) and Kitasato University Oriental Medicine Research Center in Tokyo, continued to research it.

Kampo medicine is now enjoying renewed public support. In 2009, a government task force suggested removing Kampo from the national insurance coverage in an effort to cap mounting medical expenses. In only three weeks, the Japan Society for Oriental Medicine in Tokyo collected nearly one million signatures from people opposed to the plan, which was abandoned.

UNIQUE FEATURES

Kampo medicine shares many similarities with TCM: they both support the concept of a gradual improvement in the body's condition using natural agents, and diagnosis is made using a pattern of symptoms. But over centuries of independent development, the two systems have diverged in character and practice.

Japan imports 80% of Kampo ingredients from China, although some herbs are grown locally. As with grapes cultivated for wine, local envi-

"If current Chinese medicine became the standard under international agreement, Kampo medicines could be wiped out."

ronmental conditions can influence the constituents of a medicinal plant. This means that any given traditional medicine, if it contains herbs grown in different places (even if they are the same species), can have variable properties. Consequently, explains Masatomo Sakurai, a researcher in traditional medicine at Kitasato University, the potency of some Kampo medicines such as kakkonto, a treatment for the common cold, can be one-tenth that of the TCM equivalent — in this case gegendang.

Cultural and political differences also separate Kampo from other Asian medicines. In China, TCM practitioners prepare their own mixtures tailored to each patient (although principal

ingredients may be standardized), and can choose whether to train in modern medicine or TCM. In Japan, by contrast, Kampo formulae are manufactured according to rules issued by the health ministry in 1987 to ensure consistent preparations. Kampo doctors must also hold a licence to practise modern medicine. "These features give Kampo an advantage when integrating with Western medicine," says Ikuo Saiki, director of the Institute of Natural Medicine at the University of Toyama. "Products of standard quality are essential to produce persuasive clinical data."

TALKING POLITICS

Kampo-related issues have recently taken on a political flavour. China has started to regulate exports of raw herbal materials, and has also raised prices. Although cultivation of some plants can be relocated to Japan, certain high-quality herbs, such as *Ephedra sinica*, can be harvested only in China with the right soil and climate.

At meetings of the World Health Organization and the International Organization for Standardization, the Chinese contingent is promoting TCM as the basis for standards in traditional herbal medicine. "If current Chinese medicine became the standard under international agreement, it would have an impact on our education and qualification systems," says Katsutoshi Terasawa, director of the Kampo department at the Chiba Central Medical Center in Japan. "Kampo medicines could be wiped out."

International standards would help avoid confusion over nomenclature and inform a scientific evaluation of traditional medicines. But "if we are to create international standards, we should first set the theme as traditional East Asian medicine, not just Chinese medicine," says Terasawa. Toshihiko Hanawa, director-general of Kitasato University Oriental Medicine Research Center, adds: "And for Japan, it is imperative that we establish a government support system." Without such support, there is a risk that TCM will predominate and a vital bridge between East Asian and Western medicine will be lost. ■

Ichiko Fuyuno is a freelance writer based in Munich, Germany.

PERSPECTIVE



Herbal dangers

Traditional plant-based remedies are not risk-free. Doctors and patients need to be informed about the possible side effects, says **Masatomo Sakurai**.

Many modern medicines come with warnings of side effects, and some patients are understandably wary of their use. There is a trend for people to consider herbal medicines as natural and therefore 'healthier' than synthetic chemicals. In Japan, doctors are increasingly prescribing herbal medicines while more and more scientists investigate their efficacy.

Although herbal remedies work gently and typically cause milder side effects than modern medicines, they can still be dangerous. Adverse events can vary from the minor (gastric distress and appetite loss) to the serious (interstitial pneumonia or renal cancer), and even death. According to the Japanese Ministry of Welfare, in just two years from 1994 to 1996, herbal medicines caused at least ten deaths in Japan. The problem of herbal side effects is common elsewhere in Asia but is overlooked.

A SEVERE PROBLEM

In Japan, the first reported side effect resulting from the ingestion of herbal medicines was in 1989, from a traditional Japanese formula called syo-saiko-to prescribed to promote liver health (xiao-chai-hu-tang in Chinese)^{1,2}. Over the following 11 years, 35 cases of side effects related to syo-saiko-to were reported, including eight deaths. Because most of these cases involved hospitals or doctors that did not specialize in herbal medicine, it was generally believed that the deaths were attributable to the prescribers' lack of knowledge.

However, recent reports reveal that herbal formulae can cause side effects even when prescribed by experienced traditional medicine practitioners. Although no deaths linked to herbal formulae have been reported since 2000, there have been other side effects. In the seven years to 2009, the Center for Japanese Traditional Herbal Medicine at Kashima Rosai Hospital in Kamisu reported 503 cases of side effects out of 2,530 total prescriptions, with symptoms varying from poor appetite to 26 cases of liver disorder — half of which were caused by herbal medicine (4 in 1,000 cases)³. The Department of Herbal Medicine at Kitasato University in Tokyo has reported a similar rate: 21 cases of liver damage out of 20,271 prescriptions over 9 years (approximately 1 in 1,000)⁴.

Across Japan, the most common side effects from herbal formulae are gastrointestinal distress, often caused by *Rehmannia* root and *Ephedra* herb, and allergic reactions such as skin rashes, which can be caused by many different herbs. More serious side effects include interstitial pneumonia, which affects 1 in 25,000 patients taking any herbal remedy⁴, most commonly with syo-saiko-to, which contains *Scutellaria* root. This root, among others, can also cause liver dysfunction — a condition that affects 1–5 people out of every 10,000 taking herbal medicines, most notably women more than 50 years of age. Renal dysfunction and cancer of the urinary system can result from the consumption of aristolochic acid — a constituent of *Aristolochia manshuriensis* that is used in guan-mu-tong⁵.

Other common herbal formulae have also produced unwanted results. Up to 3% of people who consume liquorice, which is added to three-quarters of Japanese herbal formulae, develop pseudohyperaldosteronism, a condition that includes hypokalaemia (low potassium levels), convulsions, oedema, hypertension and weight gain. Medicines containing the root of *Bupleurum* can give someone an urgent need to urinate, a feeling of incomplete urination, and blood in the urine. These are also the symptoms of cystitis, a form of bacterial bladder infection. The *Bupleurum*-related cases, however, are thought to be an allergic reaction to the herb, as there are no signs of bacterial infection.

Side effects can be prevented to some extent if doctors are properly trained in the use of herbs and herbal formulae. For example, the *Ephedra* herb contains ephedrine — a stimulant often used in herbal medicine that should not be prescribed to patients with cardiovascular diseases.

Unfortunately, not all side effects can be predicted, even if they are prescribed by well-trained doctors. Here, however, is some simple and practical advice that could help prevent serious cases of side effects from occurring.

First, use a blood test to check liver and renal function before the first prescription, and every 3 months thereafter. This testing will enable early detection of dysfunction, which might not manifest before 12 months. Second, measure a patient's blood pressure at least once every 2 weeks. This helps to detect pseudohyperaldosteronism. Third, inform the patient about the early symptoms of the treatment's side effects, such as fever, oedema and shortness of breath, which might develop into a more serious condition. Advise the patient that if these symptoms occur, they should cease taking the herbal medicine and inform the prescribing doctor.

Building the education system is an urgent matter for countries looking to improve the safety and efficacy of herbal formulae. Japan has now made

it compulsory for medical students to take courses in the practice of herbal medicine, and in 2008 the first doctors graduated from this system. More importantly, doctors, researchers and government officials in Japan, China and South Korea should work together to decide the common name for each herb and create international rules concerning the use of herbal formulae to improve the sharing of information about side effects.

Masatomo Sakurai is a researcher of oriental medicine, acupuncture and herbal medicine at Kitasato University Oriental Medicine Research Center in Tokyo, Japan.
e-mail: sakuraim@insti.kitasato-u.ac.jp

1. Ito, T. et al. *Kampo Med.* **61**, 299–307 (2010).
2. Gono, Y. *Kampo Med.* **61**, 828–833 (2010).
3. Arai, S. *Kampouigaku* 124–130 (2009).
4. Vanherweghem, J. L. et al. *Lancet* **341**, 387–391 (1993).
5. Nortier, J. L. et al. *N. Engl. J. Med.* **342**, 1686–1692 (2000).



Aconite root is used to treat colds and can cause palpitations and blood rushes.



JOHN GREIM/SCIENCE PHOTO LIBRARY

REGULATIONS

Herbal medicine rule book

Can Western guidelines govern Eastern herbal traditions?

BY NATASHA GILBERT

In the early 1990s, thousands of women attending a slimming clinic in Belgium were accidentally given a weight-loss treatment containing the toxic herb *Aristolochia fangchi* rather than the anti-inflammatory agent *Stephania tetrandra*. For practitioners and adherents of herbal medicine, it was one of their worst nightmares: more than 100 of the women suffered kidney failure. Many later developed cancer of the urinary system.

This was partly a story of mistaken identity: Chinese herbs are traded using their common names, which can confuse Western practitioners of traditional medicine. In this instance, the term fang ji describes the roots of both *A. fangchi* and *S. tetrandra*. In addition, in traditional Chinese medicine, *Aristolochia* species are regularly interchanged with other plant species, further complicating the use of these herbs for Western doctors who may be unfamiliar with the language and with traditional practice.

The Belgian cases prompted the European Medicines Agency to create new regulations, which are considered to be some of the strictest in the world. Given the increasing popularity among Western populations of herbal medicinal products — around a third of UK adults use

herbal products, for example — there is a pressing need to tackle safety issues while ensuring the availability of potentially useful therapies to patients.

But are Western medical agencies capable of regulating herbal medicines developed in Eastern traditions? And if not, what reforms are needed to make it possible?

NEW EUROPEAN DIRECTIVE

The Traditional Herbal Medicinal Products Directive (THMPD) came into force across the European Union (EU) in April 2011. The THMPD “aims to protect public health and at the same time secure the free movement of herbal products within the EU”, according to the European Medicines Agency.

Under the directive, herbal medicines intended as treatments for minor health ailments must be registered as traditional-use products with the regulatory agency in every EU member state in which the product is to be sold (although each national regulatory agency is supposed to recognize licences already granted by other EU members). This system is designed to make it less likely that a product is sold as a traditional herbal medicine in one country and as something else in another. Before the THMPD, Europe had a patchwork

of herbal medicine regulations. Some countries did not regulate traditional herbs at all; others classified herbal products as food supplements rather than medicines, and so subjected them to less scrutiny. Consequently, some traditional herbal products had been available to patients without the quality and safety guarantees that come with registration.

With the THMPD, criteria for whether a product is a medicine or a food supplement are consistent across all European member states. Moreover, these criteria apply to all manufactured herbal products whether they are sold in healthfood shops or pharmacies, or prescribed by doctors (as happens in Germany). The THMPD does not cover practitioners of herbal medicine, who are still permitted to mix herbal remedies from individual components for personal prescriptions.

Products are eligible for licence as a traditional herbal medicine only if they have been used to treat a specified health complaint for at least 30 years, including a minimum of 15 years in Europe. Herbal products are held to similar safety and quality standards as pharmaceutical drugs. What's different is how manufacturers are required to demonstrate efficacy. Rather than having to conduct original clinical trials, as pharmaceutical drug manufacturers must,

makers of traditional herbal medicines are instead permitted to point to their long history of use. “Some of these medicines have been around for thousands of years,” explains Dick Middleton, technical director at Schwabe Pharma UK in Buckinghamshire. “If they didn’t work they would have disappeared by now.”

WEIGHING UP THE RULES

Many Western herbal providers and manufacturers applaud the move to improve standards, and expect Europe’s new rules will provide a fairer ground for competition among responsible herbal suppliers.

“We are supportive of the rules because they ensure herbal products are consistently of an acceptable standard and give patients accurate information,” says Middleton. Schwabe Pharma UK — part of Dr Willmar Schwabe Pharmaceuticals, a German phytomedicine company based in Karlsruhe — has registered 18 traditional herbal medicine products with the UK’s Medicines and Healthcare Products Regulatory Agency — the most by any company.

Other supporters of herbal medicine worry that the rules are too demanding and could ban some therapies on which people depend. Adam Smith, science and communications officer for the Alliance for Natural Health International (ANHI) in Dorking, United Kingdom, a non-governmental campaign group promoting the use of herbal medicines and other approaches to healthcare, fears that patients will lose out on some Asian medicines because they have not been used in Europe for the requisite 15 years, even though they have been consumed in East Asia for considerably longer.

ANHI also detects a perceived bias in the THMPD towards products developed in the West, which often contain just one herb. Traditional Asian products contain mixtures of several herbs — making it difficult, time consuming and expensive to meet the directive’s requirement to identify and quantify the active botanical ingredients or other biological agents in a herbal product. “These technical assessments require expensive methods,” says Smith. “The cost burden is a problem, particularly for small businesses,” he adds.

Middleton agrees that conducting scientific analyses are “tough” — even for products containing just one herb, such as echinacea (for colds), St John’s wort (depression and anxiety) and feverfew (migraines) — all registered under the THMPD. Attempting to determine the biochemical characteristics of the ingredients in mixtures of up to 14 herbs sometimes found in traditional Chinese medicines can be an “endless story,” says Rob Verpoorte, a pharmacologist and molecular biologist at Leiden University in the Netherlands, who studies medicinal plants. This biochemical requirement is a likely reason that, of the 350 herbal medicines that had been licensed under the THMPD by December 2010, none were from an East Asian background. The European Medicines Agency says

it predicts a “substantial increase” in registrations in an updated assessment due in December 2011, but again few of these are expected to be from East Asian traditions.

Removing all but essential ingredients could simplify the analysis, suggests Arnold Vlietinck, a pharmaceutical scientist at the University of Antwerp, Belgium, and chairman of the regulatory affairs committee of the Society for Medicinal Plant and Natural Product Research, based in Bonn, Germany. For example, he says, other herbs are sometimes added to preparations to improve the look and taste.

In the longer term, ANHI suggests dispensing with the THMPD entirely and developing a new European regulatory framework for traditional medicinal products that is distinct from the existing regimens for food and medicine. This new framework would deal with all medicinal herbal preparations — including over-the-counter products and those prescribed by practitioners, as well as products on sale in some countries as food supplements.

AMERICAN DIET

In the United States, the regulatory requirements for herbal medicinal products are simpler than in Europe. US herbal products are generally regulated as dietary supplements, meaning that standards are lighter. Manufacturers do not need to analyse the biological and chemical properties of their herbal products, and there is no need to register them with the Food and Drug Administration (FDA). This laissez-faire regulation only goes so far, though. Sellers of herbs in the US are permitted to make only limited health assertions. In particular, US regulations prohibit the claim that any herbal product prevents disease. Moreover, US manufacturers are responsible for making sure that the information on the herbal product label is truthful and not misleading, which the FDA monitors — alongside reports of adverse effects from the use of dietary supplements.

This simpler system is not without problems, says Darrell Rogers, communications director at ANHI’s US branch, based in Washington, DC. Herbal products formulated to treat more serious ailments — a category that covers many traditional Asian preparations — fall between the regulatory cracks. If products are marketed as food supplements, their makers cannot make claims about the conditions they aim to treat — and people suffering from particular maladies will be unaware of these potentially beneficial treatments. To make therapeutic claims, the products have to register as medicinal drugs and undergo the FDA’s rigorous, expensive and lengthy drug approval process — which, like the European regulations, requires complex biochemical analyses and three phases of clinical trials. It’s not surprising that only one herbal medicine has achieved this status: sincatechins, a tea polyphenol for the topical treatment of cervical warts. But these are something of a special case, and are registered as a ‘botanical’ — a separate classification from synthetic, chemically pure drugs.

One pharmacologist attempting to get an Asian medicine approved by the FDA is Yung-Chi Cheng at Yale University in New Haven, Connecticut. For the past 11 years, Cheng has been investigating PHY906, a 1,800-year-old Chinese formulation of four herbs that has the potential to alleviate the vomiting and cramps caused by chemotherapy for gastrointestinal cancer. With the latest analytical technologies available to him, Cheng has made good progress: he has completed phase I clinical trials of the herb and is soon to start phase II. Cheng is investigating the concoction’s effects on three tissue types — splenic, hepatic and the gastrointestinal tumour tissue — to see if it works differently. Cheng’s research has revealed some surprises: the herbal mixture not only relieves sickness during chemotherapy but also seems to enhance the patient’s response to treatment¹. “When the chemical goes into the body it doesn’t only act on one organ, it also affects others and might help the final outcome of the patient,” he adds.

Cheng says his experience highlights the need for more research into traditional Asian medicines — and for a change in attitude, both from regulators and some scientists. He had a lot of negative feedback about his compound, he says, before he finally got it published. “The mainstream science journals take one look at it, see it is a mixture and reject it,” he says.

Work like Cheng’s shows that traditional Asian medicines could provide important new avenues for treatment. But without an appropriate regulatory framework, these could be lost to science. Getting Eastern and Western medical traditions to meet will require political and cultural changes as much as scientific ones. ■

Natasha Gilbert is a reporter for *Nature* in London.

1. Lam, W. *et al. Sci. Transl. Med.* **2**, 45ra59 (2010).



PERSPECTIVE



Liang Liu

The clinical trial barriers

To investigate traditional Asian medicines properly, we need to rethink the way they are tested, say **Liang Liu, Elaine Lai-Han Leung and Xiaoying Tian.**

Western medicine depends on science to create and assess drugs at the molecular level. In Asia, there is a commonly held belief that there is an art to healing too, and that both art and science should cooperate to help eradicate illness and relieve suffering.

Part of this art is seen in the way that Asian doctors perform a diagnosis. Doctors identify multiple signs and symptoms across the whole body and then consider these as a constellation or pattern that is unique to each person. Prescription, therapy and dosages are tailored to each patient and are often modified as the course of treatment progresses. Such complexity and changeability has been difficult, if not impossible, to assess through the current standard method of randomized controlled clinical trials (RCTs). Therefore, there is a huge need to design a rigorous yet appropriate clinical trial that can assess the safety and efficacy of traditional Asian medicines.

RCTs are the gold standard for testing new medicines. Within an RCT, a person will be randomly allocated to a group that receives one of the treatment options or a placebo. Typically, each group should contain the same distribution of relevant demographics, such as age, sex and ethnicity, among others.

RCTs can be classified into two subtypes depending on what they seek to prove. Explanatory RCTs (ERCTs) evaluate the efficacy of a pharmaceutical agent under highly controlled conditions that do not necessarily reflect real life. Pragmatic RCTs (PRCTs) assess the effectiveness of a treatment in everyday situations, meaning that a less-than-strict adherence to the treatment regimen is not a reason to exclude a subject. Other than that, both subtypes of RCT share the same principles: randomization of subjects into treatment and control groups; blinding all participants as to which group subjects are in; the need for follow up; and the handling of drop-outs and withdrawals¹. ERCTs can produce more precise assessments of a pharmaceutical agent, although the trial design often deviates significantly from daily routine. PRCTs, on the other hand, generate weaker evidence but better reflect the circumstances of patients as they go about their daily lives.

The rigidity of ERCTs has necessitated the concept of the 'average patient'. Yet such a patient does not really exist, and designing a trial in this manner is not ideal for tailoring treatment to an individual. Without prior consideration of the underlying genetics, patients in a particular trial group might receive a drug that is ineffective for their genotype — or worse still, dangerous¹. With the growing appreciation that a person's genetic variation affects response to treatment, trial designs must adapt.

Knowledge that dysregulated genes and proteins underlie the pathogenesis of many diseases has led to the development of pharmacogenomics: targeted molecular or genotype-based therapies that improve treatment and reduce adverse effects. However, RCTs have struggled to keep pace with the emerging understanding of individual pharmacogenomics². For example, gefitinib, a lung-cancer drug, inhibits epidermal growth factor receptor (EGFR) signalling in target cells, providing patients with a significantly longer period without disease progression and fewer side

effects than with conventional chemotherapy³. Only patients with a specific mutation in the gene that encodes EGFR get any benefit from gefitinib. Yet initial clinical trial results were negative because none of the patients in the trials had been selected based on EGFR activity. Positive clinical results were seen only after these subgroups were identified^{4,5}.

Integrating the principles of personalized medicine into RCTs leads to personalized explanatory RCTs (PERCTs) and personalized pragmatic RCTs (PPRCTs). These types of trials select patients according to their genotype, rather than the normal methods based on shared clinical diagnostic and biological characteristics. Moreover, this concept of personalized medicine has similarities with the individualized diagnostic and treatment methods of traditional Asian medicines (notwithstanding the fact that the molecular mechanisms have yet to be elucidated). Therefore, it is reasonable to assume that any clinical trial designed for

personalized medicine should be adaptable for testing traditional Asian medicines.

The diagnostic principles of Asian and Western medicine can be bridged by systems biology, which is a trend in biomedical research to examine complex systems in their entirety, rather than take a more reductionist, molecular viewpoint. Work is ongoing to link traditional Asian diagnoses and treatments with proteomic and biochemical signatures (see 'All systems go', page S87), with the promise that these can be applied in PERCTs and PPRCTs, as part of the criteria for assessing treatments⁶. For example, a recent report

showed that the traditional Asian pattern diagnosis of 'heat' or 'cold' in rheumatoid arthritis patients was associated with different underlying genomic and metabolomic profiles, and different treatment outcomes⁷.

For traditional Asian approaches to become integrated into modern medicine, we need to consider alternative inclusion and exclusion criteria for RCTs. We also need to review how outcomes are assessed. These concepts are neither new nor unproven. Both Western-style modern medicine and traditional Asian medicine aim to heal patients in a harmonized way and can be developed together into an integrated form of personalized medicine. Redesigning clinical trials will accelerate the blending of these two styles of healing, for the benefit of humankind. ■

Liang Liu is vice-rector at State Key Laboratory for Quality Research in Chinese Medicines, Macau University of Science and Technology, China. **Elaine Lai-Han Leung** is a researcher at State Key Laboratory for Quality Research in Chinese Medicines, Macau University of Science and Technology, China. **Xiaoying Tian** is a researcher at the School of Chinese Medicine, Hong Kong Baptist University. e-mail: lliu@must.edu.mo

- Swanton, C. & Caldas, C. *Ann. N. Y. Acad. Sci.* **1210**, 34–44 (2010).
- Frueh, F. W. *Pharmacogenomics* **10**, 1077–1081 (2009).
- Douillard, J. Y. et al. *J. Clin. Oncol.* **28**, 744–752 (2010).
- Lynch, T. J. et al. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
- Paez, J. G. et al. *Science* **304**, 1497–1500 (2004).
- van der Greef, J. et al. *Planta Med.* **76**, 2036–2047 (2010).
- van Wietmarschen, H. et al. *J. Clin. Rheumatol.* **15**, 330–337 (2009).



BIODIVERSITY

Endangered and in demand

With an ingredients list that includes rhino horn and tiger bone, traditional Asian medicine is on a collision course with wildlife preservation.

BY DUNCAN GRAHAM-ROWE

It looks innocuous enough: a small vial bearing a white and orange label with the words 'Shi-He Ming Yan Wan'. Yet the pills contained within are said to hold great healing powers, able to cure just about anything, from a mild fever to a brain haemorrhage; from cancer to AIDS. The pill's power, it is claimed, comes from a small amount of rhinoceros horn. Little wonder then that people pay as much as US\$50,000 for a kilogram of the stuff, roughly the same as the price of gold (see 'The rhino toll').

The rhino and its horn are not alone: powdered tiger bone is used to treat rheumatism; the scales of the toothless, anteater-like pangolin are believed to reduce swelling and improve blood circulation; and *guilinggao*, a jelly derived from the shells of freshwater turtles, was used to treat smallpox in a nineteenth-century emperor, with little success — in Taiwan it is now reputed to cure cancer. It is a similar story for many other endangered species whose commercial use is restricted — or banned outright — by the Convention on International Trade of Endangered Species of Wild Fauna and Flora (CITES).

The illicit trade in wildlife is a booming industry, estimated by the US congressional research service to be worth as much as US\$20 billion globally each year¹. Although this figure includes trade in bushmeat, skins and exotic pets, in the expanding Asian market, estimated to be the largest in the world, a significant driver is traditional Chinese medicine (TCM). Indeed, despite showing signs of decline in the 1990s, the poaching and trade of endangered animals such as tigers and rhinos is once again on the rise. Yet cheaper and more potent alternatives are available. Organizations such as the American College of Traditional Chinese Medicine say that sustainable substitutes have been used successfully for nearly two decades. So why is there still a burgeoning market to use these precious animals in traditional Asian medicine?

RICH PERSON'S REMEDY

One likely factor driving this demand is the rise in the wealth of China, says Sabri Zain, director of advocacy for Traffic International in Cambridge, UK, which was established in 1973 to monitor wildlife trade. "Currently China is the biggest market," he says. This dominance is not just a consequence of China's population, or the fact that traditional Asian medicine has its roots there, but to the country's rapidly rising incomes. "There are more people who can afford it," Zain says.

The market for these substances also seems to be expanding. A range of new products has emerged over the past decade, available as black market products or through online stores. "Tiger bone is now being used in wine," says Debbie Banks, a senior tiger investigator with the Environmental Investigation Agency (EIA), a campaign group in London. It is a similar story with other new products made with tiger bones, such as shampoo — or indeed with tiger penis

soup, which has no perceived medicinal value whatsoever, she says. In fact, these wines, sham-poops and soups are not part of the traditional medicine repertoire at all. However, they do lean on the same beliefs, says Banks. “They are seen as status products,” she adds.

Indeed this association with status is a major issue, reflected in the demographic of the modern-day user of these products, says Zain. “It’s a myth that these products are only being consumed by an older generation. It’s not just the old fogies, it is also young, wealthy professionals,” he says. “It may be a way of showing their peers that they can afford these very expensive medicinal products.”

And there appear to be plenty who meet that financial test. In 2008, a survey of nearly 1,000 people from six cities across China found that 1.9% of respondents had consumed a medicine or tonic containing tiger within the past 12 months². If this represents national consumption, it would mean a user base of around 25 million people.

In Vietnam, which is one of the largest markets for TCM outside China, traditional remedies are sought after. If incomes were to increase, so too would consumption of products containing endangered species³. This is hardly surprising, says Zain, given the perception that products such as rhino horn are capable of curing cancer — a medicinal property previously unheard of in traditional Asian medicine.

HORN OF PLENTY

So what of these purported health benefits? The use of rhino horn dates back at least 1,800 years and is referenced in Shennong’s *Classic of Materia Medica*, the very first book of Chinese herbal medicine, says Huijun Shen, president of the UK Association of Traditional Chinese Medicine in London. Shen has nearly 30 years experience as a traditional medicine practitioner, having trained and practised in China before moving to the UK in 1995. Tiger bone is a more recent addition to the TCM apothecary but is similarly well established in the Chinese literature as a medicinal treatment. Yet despite their place in the East Asian medicinal tradition, the clinical applications of both rhino horn and tiger bone are limited, so their use is rare. Even before tiger and rhino products were banned in China in 1993, demand for them was weak, says Shen. “I never used them, even when they were legal,” he recalls, “because they were so expensive.”

In fact, in nearly two millennia, there is no record of rhino horn as a treatment for cancer, says Shen. Rhino has very few medicinal uses, the main one being to treat what is known in TCM as ‘heat toxin in blood’ syndrome (there is no direct modern medicine equivalent, although it is usually caused by serious infections and can lead to toxæmia or septicaemia). Today this condition is treated with cheaper and more effective alternatives. “Normally we just use antibiotics,” he says. Modern TCM practitioners, he adds, are often

trained in both TCM and Western medicine.

According to Zain, anecdotal evidence suggests that using rhino horns to treat cancer was fabricated recently by illegal traders to boost their market. And, judging by the increase in poaching, it appears to be working. “Rhino is probably the most worrying trend,” says Zain. That’s because, until 2007, the number of rhinos being poached seemed to be dropping dramatically. Since then, however, there has been a resurgence, from 8–10 illegal rhino horn seizures a year to 200 or more in 2007. It’s not entirely clear what has caused this increase, but seizures on the ground and at international borders are still on the rise — this year the tally is expected to reach more than 400. Moreover, Zain says, based on the number of poached rhinos, the number of illegally traded horns that evade capture by the authorities is also increasing (see ‘The rhino toll’). Indeed, in autumn 2011, the International Union for the Conservation of Nature announced that wild populations of black rhino in West Africa are considered officially extinct.

It is a similar story with tigers, says Banks. Back in 1993, after intense international pressure and under the threat of sanctions from the US government, China’s state council prohibited the import or export of tiger bone and rhino



Tiger-bone remedy for sale in Vietnam.

horn, and banned the sale of medicinal products containing them within the country (with an exception for research into substitutes) under penalty of death. In 1999, China established an anti-smuggling unit. Following these measures, according to Chinese CITES officials, seizures of tiger parts seemed to decline, as did the availability of tiger products on the street⁴.

INVESTING IN EXTINCTION

Since 2002 the declining trend has reversed, and seizures of tiger parts are now four times what they were a decade ago. Tiger skins alone can fetch in excess of US\$20,000 per hide, so this rise cannot be attributed solely to supplying traditional medicine. But with each set of tiger bones worth as much as \$7,000 it is unlikely that a smuggler would sell the skin without also trying to profit from the bones.

Even if ever more affluent consumers in

countries such as China and Vietnam are ramping up demand for traditional Asian medicines, why opt for hugely expensive products, and potentially risk the death penalty, when alternatives are available that are not only cheaper, but which are also claimed to be more effective? China’s Northwest Institute of Biology has conducted many different studies into the properties of *sailong* (mole rat) bone and its use in treating rheumatism. Moreover, there are a number of drugs containing it approved by the Chinese regulatory agency to treat rheumatism. The Chinese authorities actively encourage using *sailong* bone as an alternative to tiger bone.

Yet many believe that the Chinese authorities could be doing much more. China does have harsh punishments for those caught trading in illegal wildlife, and regular seizures of animal contraband are made at its borders. But within China there is little policing, says Banks; bones and skins are openly sold in markets. And, as of 2011, the death penalty is no longer applied to this crime.

What’s more, the Chinese government’s attitude to farming animals for use in traditional medicine is a concern, says Banks. Bears are farmed for their bile, which is used to treat gallstones. There are also many officially sanctioned tiger farms⁴. Most only hold a dozen or so tigers, says Alasdair Cameron, a campaigner with the EIA, but several house more than a thousand of the big cats. He estimates that there are between 6,000 and 10,000 captive tigers in the country. And with reports of China importing hundreds of live rhinos too, it raises the question of why they are breeding these animals if not for their bones and skin.

“The Chinese government has categorically denied that this is what they are doing — and that these live rhinos have been brought into the country to meet the demand for recreational zoos,” says Zain. But the cramped conditions the animals are kept in suggests otherwise, he says. More to the point, some of these tiger farms were set up before 1993 with the stated aim of breeding tigers for use in traditional medicine and for their skins — before the ban on tiger bone was introduced — and have never closed, says Cameron.

“You could argue that these breeding farms will take the pressure off wild animals,” says Zain. This is the argument for crocodile farms, which produce meat and skin. But the reality is that farms maintain demand for the animal parts, yet cannot cope with any surge in demand. If this happens, the temptation will be to satisfy it by hunting — putting pressure on wild populations. “There are millions of crocodiles in the world,” Zain observes, “but only a few thousand tigers.” So farming tigers not only encourages the capture of wild animals for breeding but also stimulates demand for the products. With so few tigers left in the wild, it would only take a small increase in demand to push them over the edge into extinction, he says. What’s more, it’s not really in the interest of tiger farmers to prevent the extinction of wild populations: after all, once

THE RHINO TOLL

An upsurge in the cost of rhino horn has fuelled poaching. Increases have coincided with new medical claims about the use of rhino horn in traditional Asian medicines.

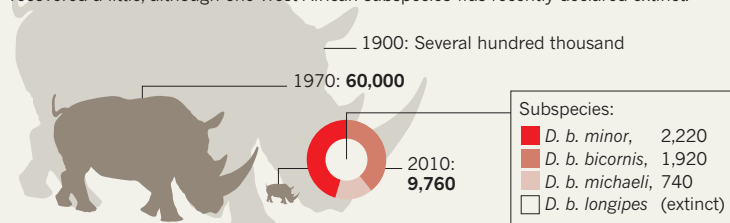
THE COST

Black rhino numbers have shrunk by 96%, primarily due to poaching.



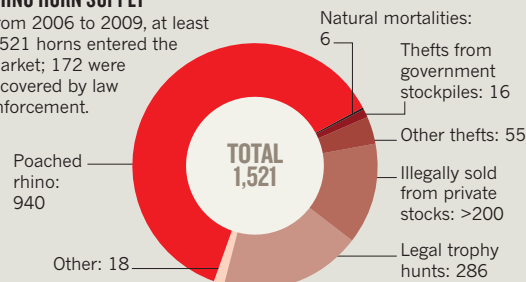
BLACK RHINO NUMBERS

From 1970, large-scale poaching of the black rhino (*Diceros bicornis*) in Africa caused numbers to fall by 96% to an estimated minimum of 2,410. Numbers have since recovered a little, although one West African subspecies was recently declared extinct.



RHINO HORN SUPPLY

From 2006 to 2009, at least 1,521 horns entered the market; 172 were recovered by law enforcement.



SEIZURES OF RHINO HORN

Many more rhino horns now evade law enforcement.



they are extinct there would be no reason to control the trade in tiger body parts.

But perhaps the most disturbing notion is the prospect that people might trade in endangered animals as a means of “investing in extinction”. This is the idea that by actively buying up and stockpiling rare animal parts, one can not only push up the price, but also encourage further poaching that will eventually force the species into extinction. In cold-blooded business terms it makes an awful lot of sense, says John Scanlon, secretary-general of CITES in Geneva, Switzerland. “If something is rare it becomes more attractive,” he says. “And the rarer something is, the more valuable it becomes.”

Scanlon concedes that he only has “rumour and anecdotal evidence” that anyone is actually ‘investing’ in the demise of a species. “It’s still speculative,” he adds. According to the EIA, however, tiger farms are stockpiling the bones and skins of tigers that die. Indeed, Chinese authorities have set up two operations — one in Guangxi and one in Geilongjiang — to dismember the carcasses of dead tigers and destroy all but the bones and skins⁴. The Chinese authorities say this is to ensure there is adequate supervision of the carcass and body parts, but why the bones and skin are then not destroyed is not clear. So although there is no proof of people stockpiling wild tiger parts, it is certainly happening in farms.

DAMPENING DEMAND

There is no simple solution to tackle illegal trade in these endangered animals, says Scanlon. But the hope is that progress can be made by adopting diverse tactics, including controlled delivery, tracking illicit substances to the buyer. “It has

been used to combat illicit trade in drugs and we are now training wildlife enforcement officials in the same technique,” says Scanlon. “We have employed two staff who have vast front-line experience with police and customs in China and South Africa to assist us in the fight against wildlife crime.” And CITES has scheduled a December 2011 training course on controlled delivery techniques for staff in 20 states in China.

Furthermore, in 2008, despite having a remit limited to international trade, CITES took the

“The rarer something is, the more valuable it becomes.”

unusual step of issuing an advisory notification to signatories, which include both China and Vietnam. The notification stated that countries in which intensive breeding of rhinos or tigers takes place should ensure that none of the animals are bred for their parts or derivatives. It even suggests the use of DNA profiling, so that illicit trade can be traced to source. But, as Scanlon points out, the notification is merely advisory and not binding.

Ultimately, Zain believes that the solution must come through demand reduction — “to make it socially unacceptable, or dispel the myths concerning these products”, he explains. “If there is no demand for a product then there is no trade.” Such an approach has already successfully reduced the demand for tiger skins among Tibetans. Following a campaign to raise awareness of the tiger’s plight, and an appeal by the Dalai Lama in 2006 for Tibetans to stop wearing the fashionable tiger-skin chubas, demand fell dramatically. Tiger skins were burned and it became socially unacceptable to wear or sell them,

and demand in Tibet has all but disappeared⁵.

The truth is that simply prohibiting trade or the use of animal parts doesn’t stop the poaching. “After the ban, demand still didn’t disappear completely,” says Shen. Even changing the traditional medicine formulae has minimal impact. “Most practitioners would not use these substances,” Shen says. But people don’t always go through a practitioner. “For TCM you don’t need a prescription — you can buy any Chinese medicine over the counter.”

Demand reduction will take time. “Cultural and historical beliefs play a big role,” says Shen, and with traditional medicine, attitudes and beliefs run deep. Providing sustainable and effective substitutes for tiger bone and rhino horn is one thing, but getting people to use them and believe in them is another. Indeed, one of the findings of Traffic International’s survey in Vietnam was that there is a powerful underlying belief that medicines made of wild animals are of a higher quality than those using farmed animals or synthetic substitutes. In essence, getting users of traditional Asian medicines to change their ways is as much about changing tradition as it is about advancing medicine. ■

Duncan Graham-Rowe is a freelance science writer based in Brighton, UK.

1. Wyler, L. S. & Sheikh, P. A. Congressional Research Service (2009).
2. Wasser, R. M & Jiao, P. B. *Traffic East Asia Report* (2010).
3. Venkataraman, B. (Comps) *Traffic Southeast Asia* (2007).
4. Chunyu, S. *CITES Communication SC61 Doc. 41, Annex 2* (2011).
5. Environmental Investigation Agency Report (2011).